

SONALI PAULA MOLIN BEDIN

**METODOLOGIA PARA VALIDAÇÃO DE ONTOLOGIAS:
O CASO ORBIS_MC**

DISSERTAÇÃO DE MESTRADO

FLORIANÓPOLIS (SC)

2007

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

SONALI PAULA MOLIN BEDIN

METODOLOGIA PARA VALIDAÇÃO DE ONTOLOGIAS:
O CASO ORBIS_MC

Dissertação Mestrado apresentada ao Programa de Pós-graduação em Ciência da Informação do Centro de Ciências da Educação, da Universidade Federal de Santa Catarina (UFSC), para a obtenção do título de Mestre em Ciência da Informação.

Area de Concentração: Gestão da Informação.

Linha de Pesquisa: Fluxos de Informação.

Orientadora: Profª. Dra. Lígia Maria Arruda Café.

FLORIANÓPOLIS (SC)

2007

SONALI PAULA MOLIN BEDIN

**METODOLOGIA PARA VALIDAÇÃO DE ONTOLOGIAS:
O CASO ORBIS_MC**

ESTA DISSERTAÇÃO FOI APROVADA PARA A OBTENÇÃO DO TÍTULO DE
**MESTRE EM CIÊNCIA DA INFORMAÇÃO NO PROGRAMA DE PÓS-
GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO** DA UNIVERSIDADE FEDERAL
DE SANTA CATARINA. LINHA DE PESQUISA: “FLUXOS DE INFORMAÇÃO”.

Florianópolis, 05 de março de 2007.

Prof^ª. Dra. Miriam Figueiredo Vieira da Cunha (Coordenadora)

BANCA EXAMINADORA:

Prof^ª. Dra. Lígia Maria Arruda Café – PGCIN/UFSC (Orientadora)

Prof^ª. Dra. Marisa Brascher Basílio Medeiros – UnB/BRASILIA

Dra. Tânia Cristina D’Agostini Bueno – IJURIS/FLORIANÓPOLIS

AGRADECIMENTOS

À minha orientadora, Prof^ª. Lígia Café, pelo entusiasmo, pelas contribuições e discussões enriquecedoras que permitiram a construção deste trabalho.

Ao Carlos, companheiro de sempre, pela compreensão e apoio na minha opção pela academia, pela pesquisa científica e construção do conhecimento.

Aos meus filhos, Yan e Tess, que entenderam e aceitaram todos os momentos, contribuindo muito com sua presença sempre tão amorosa.

À minha família, pelo apoio em todos os momentos.

À equipe valorosa do IJURIS, em especial, Aline Nicolini, Filipe Costa e Thiago Silva, pela cooperação, pela ajuda, pelas discussões.

À Dra. Tânia Bueno, pela colaboração, disponibilidade e generosidade no compartilhamento do conhecimento em sua incansável dedicação à pesquisa científica.

Aos amigos Hermes, Marília e Adriane, pelo companheirismo.

Aos professores do Curso de Pós-Graduação em Ciência da Informação da UFSC, pela possibilidade de momentos tão enriquecedores.

À Cecília, secretária do PGCIN, pela atenção e presteza em todas as situações.

Aos professores do Curso de Biblioteconomia da UDESC, que mostraram a diversidade da Ciência da Informação e sempre foram grandes incentivadores.

Ao Prof^º. Divino Ribeiro Jr., que fez nascer em mim o gosto pela pesquisa e marcou o ponto de partida de toda a trajetória acadêmica até aqui.

RESUMO

BEDIN, Sonali Paula Molin. **Metodologia para validação de ontologias: o caso ORBIS_MC**. 2007. 113f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2007.

O desenvolvimento científico e tecnológico vem trazendo em seu escopo novas estruturas de conteúdos. Relações de toda ordem montam teias que ligam dados, informações, pessoas e locais em um mesmo ambiente não mais físico. A velocidade determinante deste novo século tem gerado inquietações acerca da nova situação que envolve todas as áreas do conhecimento. Grandes volumes de conteúdos precisam ser recuperados com rapidez e precisão atendendo às expectativas dos usuários. Sistemas de Recuperação de Informação são desenvolvidos na tentativa de solucionar esta questão. A pesquisa se insere no universo das linguagens aplicadas nesses sistemas e sua motivação é fruto da leitura e reflexão de estudos anteriores sobre o tema, envolvendo aspectos da sua estrutura, conceitos, relacionamentos e variações. Abordando a linguagem natural e suas novas formas de representação, são estudadas as ontologias de um Sistema de Recuperação de Informações que utiliza a tecnologia KMAI®. Esta tecnologia adota técnicas de inteligência artificial e ontologias como forma de recuperação da informação em documentos não-estruturados disponíveis em mídias escritas: jornais, revistas e sites eletrônicos com código aberto. O sistema-alvo de estudo, desenvolvido para o Observatório Regional Base de Indicadores de Sustentabilidade Metropolitano de Curitiba- ORBIS_MC, busca estruturar informações que vão subsidiar políticas públicas de desenvolvimento para a região metropolitana por meio de monitoramento, recuperação e composição da base do conhecimento sobre Saúde Ambiental, Saúde Humana e Violência Urbana na Região Metropolitana de Curitiba. O sistema conta com um dicionário de termos construído por especialistas nas áreas definidas como objeto de representação. A presente pesquisa propõe uma metodologia de validação das ontologias que integram este dicionário, considerando os critérios lingüísticos que reflitam o uso da língua no contexto comunicativo, visto que se trata de uma linguagem utilizada para recuperar conteúdos cuja função é comunicar. Apoiando-se no estudo de publicações já reconhecidas nas áreas de Ciência da Informação e Lingüística, foram sistematizadas recomendações sobre o estabelecimento das relações semânticas entre os termos, a fim de definir critérios de validação de ontologias sob o ponto de vista lingüístico. Esses critérios foram aplicados em um “*corpus*” de 28 termos e 163 relações semânticas. O resultado da análise apresenta 82,21% de relações semânticas consideradas consistentes e 17,79% inadequadas. Para este último conjunto, foram tecidas recomendações. A aplicação de uma metodologia de fundamentação lingüística sob a perspectiva do uso foi a obtenção de uma ontologia atualizada, tanto no seu aspecto lexical como semântico, aumentando a eficiência de recuperação da informação.

Palavras-chave: Ontologias. Validação de ontologias. Recuperação da Informação.

ABSTRACT

BEDIN, Sonali Paula Molin. **Methodology for ontology validation: the case ORBIS_MC**. 2007. 113 f. Dissertation (Master in Information Science Degree) – Post Graduation Program in Information Science, Federal University of Santa Catarina, Florianópolis, 2007.

The technological and scientific development comes bringing in their purpose new structures of contents. Relations of all order mount webs that link data, information, people and localities in a same environment not more physical. The determinant speed of this new century, has generated restlessness about the new situation that involves all the knowledge areas. The large number of contents needs to be recuperated with quickness and precision attending to the users' expectations. Information Retrieval systems are developed in the attempt to solve this question. The research is inserted in the universe of the languages applied in these systems and its motivation is the reading fruit and reflection of previous studies about the subject involving aspects about structure, concepts, relationships and variations. Approaching the natural language and its new forms of representation will be studied the ontologies used in an Information Retrieval System that utilizes the technology KMAI. This technology adopts techniques of artificial intelligence and ontologies as form of information retrieval in documents not structured available in written medias: periodics, magazines and sites with open code. The study system, developed for the Regional Observatory Metropolitan Sustainability Indicator Base in Curitiba- ORBIS_MC, searches to structure information that are going to subsidize public politics of development for the metropolitan region by means of gathering, recuperation and composition of knowledge base about Environmental Health, Human Health and Urban Violence in the Metropolitan Region in Curitiba. The system counts with a dictionary of terms built by specialists in the definite areas as object of representation. The present research proposes a validation methodology of the ontologies that integrate this dictionary, considering the linguistic criteria that reflect the language use in the communicative context seen that it is a matter of a language used to recuperate contents which function is to communicate. Supporting itself in the study of publications already recognized in the areas of Information Science and Linguistic, were systematized recommendations about the establishment of the semantic relations among the terms in order to define criteria of ontologies validation under the linguistic viewpoint. These criteria were applied in a "corpus" of 28 terms and 163 semantic relations. The result of the analysis presents 82,21% of semantic relations considered consistent and 17,79% inadequate. For this last joint, indicated recommendations. The application of a methodology of linguistic substantiation under the perspective of the use is the obtaining of an ontology brought up to date, so much in his lexical as semantic aspect, increasing the efficiency information retrieval.

Keywords: Ontology. Ontology evaluation. Information retrieval.

RESUMEN

BEDIN, Sonali Paula Molin. **Metodología para validación de ontologías: caso ORBIS_MC**. 2007. 113f. Disertación (*Master* en Ciencia de la Información) – Programa de Pos-Graduación en Ciencia de la Información, Universidad Federal de Santa Catarina, Florianópolis, 2007.

El desarrollo científico y tecnológico viene trayendo en su escopo nuevas estructuras de contenidos. Relaciones de todo orden arman telas que unen datos, informaciones, personas y locales en un mismo ambiente ya no más físico. La velocidad determinante de este nuevo siglo ha generado inquietudes acerca de la nueva situación que envuelve todas las áreas del conocimiento. Grandes volúmenes de contenidos precisan ser recuperados con rapidez e precisión atendiendo a las expectativas de los usuarios. Sistemas de Recuperación de Información son desarrollados en la tentativa de solucionar esta cuestión. La pesquisa se inserta en el universo de los lenguajes aplicados en estos sistemas y su motivación es fruto de la lectura y reflexión de estudios anteriores sobre el tema, envolviendo aspectos de su estructura, conceptos, relacionamientos y variaciones. Abordando el lenguaje natural y sus nuevas formas de representación, son estudiadas las ontologías de un Sistema de Recuperación de Informaciones que utiliza la tecnología KMAI®. Esta tecnología adopta técnicas de inteligencia artificial y ontologías como forma de recuperación de la información en documentos no-estructurados disponibles en medios gráficos: periódicos, revistas y *sites* electrónicos con código abierto. El sistema-objeto de estudio, desarrollado para el Observatorio Regional Base de Indicadores de Sustentabilidad Metropolitano de Curitiba-ORBIS_MC, busca estructurar informaciones que van a subsidiar políticas públicas de desarrollo para la región metropolitana por medio del acompañamiento, recuperación y composición de la base del conocimiento sobre Salud Ambiental, Salud Humana y Violencia Urbana en la Región Metropolitana de Curitiba. El sistema cuenta con un diccionario de términos construido por especialistas en las áreas definidas como objeto de representación. La presente pesquisa propone una metodología de validación de las ontologías que integran este diccionario, considerando los criterios lingüísticos que reflejen el uso de la lengua en el contexto comunicativo, visto que se trata de un lenguaje utilizado para recuperar contenidos cuya función es comunicar. Apoyándose en el estudio de publicaciones ya reconocidas en las áreas de Ciencia de la Información y Lingüística, fueron sistematizadas recomendaciones sobre el establecimiento de las relaciones semánticas entre los términos a fin de definir criterios de validación de ontologías bajo el punto de vista lingüístico. Estos criterios fueron aplicados en un “corpus” de 28 términos y 163 relaciones semánticas. El resultado del análisis presenta 82,21% de relaciones semánticas consideradas consistentes e 17,79% inadecuadas. Para este último conjunto, fueron tejidas recomendaciones. La aplicación de una metodología de fundamentación lingüística bajo la perspectiva del uso es la obtención de una ontología actualizada, tanto en su aspecto lexical como semántico, aumentando la eficiencia de recuperación de la información.

Palabras-llave: Ontologías. Validación de ontologías. Recuperación de la Información.

LISTA DE FIGURAS

Figura 1: Cadeia descendente das relações gênero/espécie	87
Figura 2: Figura 2: Alteração da relação dos termos “gangues juvenis” e “gangues nas escolas”	92

LISTA DE QUADROS

Quadro 1: Domínios e subdomínios	65
Quadro 2: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Energia	66
Quadro 3: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Gestão de Recursos Naturais	67
Quadro 4: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Saneamento Ambiental	67
Quadro 5: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Poluição Ambiental	68
Quadro 6: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Ocupação do Solo	68
Quadro 7: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Órgãos vinculados à Saúde Ambiental	69
Quadro 8: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Promoção da Saúde.....	69
Quadro 9: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Vigilância em Saúde	69
Quadro 10: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Políticas de Saúde	70
Quadro 11: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Gestão de Saúde.....	70
Quadro 12: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Modalidades de Violência Urbana.....	71
Quadro 13: Universo de termos e relações analisados. Domínio: violência urbana. Subdomínio: violência baseada por localidade	71
Quadro 14: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Causas da Violência Urbana.....	71
Quadro 15: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Consequências da Violência Urbana	72
Quadro 16: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Segurança Pública.....	72
Quadro 17: Síntese de Autores e Critérios. Relações de Equivalência – Sinonímia	73

Quadro 18: Síntese de Autores e Critérios. Relações Hierárquicas – Tipo de / Parte de ...	73
Quadro 19: Síntese de Autores e Critérios. Relações Associativas – Conexos	74
Quadro 20: Definição dos conceitos que compõem os parâmetros das relações associativas apresentada por Marroni (2006)	75
Quadro 21: Síntese relação associativa causa/efeito	90
Quadro 22: Relações construídas e relações propostas	91

LISTA DE TABELAS

Tabela 1: Síntese quantitativa das relações semânticas	77
Tabela 2: Pesos para o cálculo de similaridade	78
Tabela 3: Síntese qualitativa das relações semânticas	79

LISTA DE SIGLAS

IA	Inteligência Artificial
ISO	<i>Information Standards Organization</i>
KMAI®	<i>Knowledge Management with Artificial Intelligence</i>
LC	Linguagem Controlada
LD	Linguagem Documentária
LN	Linguagem Natural
ONU	Organização das Nações Unidas
ORBIS_MC	Observatório Base de Indicadores de Sustentabilidade Metropolitano de Curitiba
RMC	Região Metropolitana de Curitiba
SRI	Sistemas de Recuperação de Informação

SUMÁRIO

CAPÍTULO 1 – INTRODUÇÃO.....	14
1.1 TEMA DE PESQUISA E JUSTIFICATIVA	15
1.2 PROBLEMA DE PESQUISA	18
1.3 OBJETIVOS	22
1.4.1 Objetivo geral.....	22
1.4.2 Objetivos Específicos	22
 CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA.....	23
2.1 TERMINOLOGIA	23
2.2 LINGUAGENS DE ESPECIALIDADE	25
2.3 UNIDADE TERMINOLÓGICA	27
2.3.1 Termo	27
2.3.2 Conceito	28
2.4 RELAÇÃO ENTRE SIGNIFICANTE E SIGNIFICADO.....	30
2.4.1 Homonímia	31
2.4.2 Polissemia	32
2.4.3 Sinonímia	33
2.5 RELAÇÕES ENTRE CONCEITOS.....	35
2.5.1 Relações de equivalência.....	36
2.5.2 Relações lógicas	36
2.5.2.1 Relação hierárquica	37
2.5.2.1a Relação genérica.....	37
2.5.2.1b Relação partitiva.....	38
2.5.3 Relação associativa	38
2.6 VARIANTES TERMINOLÓGICAS.....	40
2.7 LINGUAGENS.....	42
2.7.1 Linguagem natural.....	42
2.7.2 Linguagem documentária	43
2.8 ONTOLOGIAS	45

CAPÍTULO 3 - CONTEXTO DE PESQUISA: O SISTEMA ORBIS_MC.....	54
3.1 APRESENTAÇÃO DO ORBIS_MC	54
3.2 APRESENTAÇÃO DO SISTEMA	55
3.3 A FERRAMENTA KMAI®.....	56
3.4 RESULTADOS DO SISTEMA.....	57
3.5 A ELABORAÇÃO DO VOCABULÁRIO	57
3.6 REPRESENTAÇÃO DO CONHECIMENTO EM DOCUMENTOS NÃO- ESTRUTURADOS	58
3.7 CONSTRUÇÃO DO DICIONÁRIO DE TERMOS.....	60
 CAPÍTULO 4 – METODOLOGIA.....	 63
4.1 CARACTERIZAÇÃO DA PESQUISA.....	63
4.2 COLETA DE DADOS (PRÉ-ANÁLISE).....	65
4.3 EXPLORAÇÃO DO MATERIAL	72
 CAPÍTULO 5 - TRATAMENTO DOS RESULTADOS	 77
5.1 ANÁLISE QUANTITATIVA DAS RELAÇÕES.....	77
5.2 ANÁLISE QUALITATIVA DAS RELAÇÕES	80
5.2.1 Relações adequadas.....	80
5.2.2 Relações inconsistentes	84
5.2.2.1 Relações classificadas originalmente como Sinônimos	84
5.2.2.1a Relações de natureza associativa.....	85
5.2.2.1b Relações de natureza hierárquica	86
5.2.2.2 Relações classificadas originalmente como Hierárquicas	88
5.2.2.2a Relações de natureza sinonímica.....	88
5.2.2.2b Relações de natureza associativa.....	89
5.2.2.3 Relações classificadas originalmente como Conexas	90
 CAPÍTULO 6 - CONSIDERAÇÕES FINAIS	 93
 REFERENCIAL BIBLIOGRÁFICO.....	 97
 ANEXOS	 101

CAPÍTULO 1 - INTRODUÇÃO

No Século XXI, os estudos dedicados ao acesso à informação, particularmente aqueles que consideram a utilização dos recursos disponíveis de simultaneidade e velocidade em que essa informação está sendo produzida, têm se tornado, além de imprescindíveis, uma questão incansável para pesquisadores de diversas áreas. As inquietações acerca dessa nova situação, que envolve todas as áreas do conhecimento, têm produzido inúmeras pesquisas.

Entretanto, os estudos que objetivam identificar a recuperação de grandes volumes de conteúdos de informação nem sempre têm atendido às expectativas de buscas. Como exemplo, pode-se citar: os conteúdos produzidos e indexados em modernos sistemas informacionais que não correspondem às expectativas de acesso; e os Sistemas de Informação desenvolvidos em larga escala e dotados de complexidade que dificultam sua utilização.

Dentro desse contexto, para se indexar os conteúdos e recuperá-los de maneira satisfatória, estudos recentes descritos na literatura especializada revelam que isto envolveria práticas e caminhos que incluem sempre o binômio: homens/máquinas e usuários/sistemas. Os dois elementos, apesar de suas diferenças, são ou deveriam ser considerados importantes, complementares e interativos.

Por causa disso, a linguagem utilizada para proporcionar a recuperação de conteúdos exerce um papel fundamental, ou seja, ela é um elo entre o usuário e o sistema. Neste estudo, esse processo é entendido não como uma linguagem de programação, mas de representação do conhecimento.

Portanto, considerando os argumentos acima, esta pesquisa se insere no universo das linguagens aplicadas em Sistemas de Recuperação de Informação (SRI), e sua motivação é

fruto da leitura e reflexão de estudos e pesquisas anteriores sobre o tema apresentado, incluindo aspectos da sua estrutura, conceitos, relacionamentos e variações.

A apresentação desta Dissertação está estruturada em seis capítulos. O Capítulo 1 apresenta uma breve contextualização da temática, a justificativa, o problema e os objetivos da pesquisa. No Capítulo 2, está a Fundamentação Teórica, na qual são abordados os elementos de natureza conceitual sobre linguagens de especialidade, termos, conceitos e suas relações, linguagens de representação do conhecimento e ontologias, objeto deste estudo. O Capítulo 3 trata do contexto da pesquisa, em que se descreve detalhadamente o Sistema ORBIS_MC, seus objetivos e componentes; e a ferramenta KMAI®, suas funcionalidades, forma de representação do conhecimento e o vocabulário inserido. No Capítulo 4, são descritos os procedimentos metodológicos, a caracterização da pesquisa, os procedimentos de coleta de dados, a exploração do *corpus* selecionado e analisado e os parâmetros adotados. O Capítulo 5 apresenta os resultados da pesquisa, em que são expostas as análises quantitativa e qualitativa do *corpus* selecionado, além das proposições elaboradas com base no estudo. No Capítulo 6, são tecidas as Considerações Finais acerca da pesquisa. Ao final, são apresentadas as obras consultadas durante este estudo e os Anexos, contendo o material selecionado e comentários.

1.1 TEMA DE PESQUISA E JUSTIFICATIVA

Desde os primórdios tempos, sabe-se que a informação tem fundamental importância em qualquer sociedade. A produção e a transmissão de dados, de forma a que outros entendam e a transformem em conhecimento, têm gerado pesquisas ao longo da história.

À medida que as investigações evoluem, novas situações são criadas, fazendo nascer outras formas de interação. Dessa maneira, novos e diversos formatos de produção da

informação estão sendo disponibilizados em larga escala, tornando o conhecimento imprescindível. Os agentes dos processos de comunicação, atuantes nesse meio, dividem sua importância com outro fator fundamental: o conteúdo. Este necessita ser transmitido de forma ágil e precisa, e deve ser entendido em sua plenitude pelos possíveis usuários, para que possa ser recuperado.

A escala crescente desse processo de produção e disseminação de informação fez surgir mecanismos voltados ao registro dos dados de forma uniforme e com estrutura bem definida, com vistas a posterior recuperação e utilização dos mesmos.

Diante disso, os SRI passaram a exercer papel fundamental. Esses Sistemas têm sido alvo de pesquisas voltadas ao seu aperfeiçoamento, na busca da melhor performance que vá de encontro às necessidades do usuário já conhecido e do usuário potencial que precisam de informação. Em outras palavras, o usuário sabe onde está a informação, mas ainda não têm recursos e/ou condições técnicas para administrar este universo.

Os vários itens que compõem um SRI, estudados exhaustivamente em diversas áreas, e a linguagem utilizada nesses Sistemas, que urge ser mais próxima do discurso e do entendimento natural do usuário, despertam especial interesse na Ciência da Informação e na Ciência da Computação, fazendo emergir uma interdisciplinaridade de ações quando da sua elaboração.

No campo da recuperação da informação, diversas linguagens têm sido construídas. O ‘tesauro’¹ é um exemplo clássico. Mais recentemente, as ontologias têm se apresentado como uma alternativa na resolução de problemas semânticos prejudiciais à recuperação de informação. Sobre esse aspecto, entende-se que as Ontologias cumprem a função de

¹ Coleção de palavras agrupadas por conceitos e títulos, e não em ordem alfabética como num dicionário. Ex: tesauro de sociologia, tesauro de economia política. Arquivo contendo sinônimos que são exibidos como alternativas para uma palavra escrita de forma incorreta, durante uma verificação de ortografia (HOUAISS, 2006).

linguagem de recuperação, uma vez que elas trazem em seu escopo uma aproximação entre a construção de expressões baseadas no conhecimento e o entendimento geral.

Observou-se que a linguagem ontológica se manifesta como uma possibilidade de representação do conhecimento, de forma que ela aproxime ou indique o que o usuário está pensando na hora da busca, ou, ainda, facilite o entendimento do Sistema acerca do que o usuário pretende. O uso das ontologias se insere neste universo, como um elo de ligação entre o usuário e o Sistema com o objetivo de aumentar a eficiência na recuperação da informação. Assim, os recursos das ontologias, por serem mais flexíveis e passíveis de avaliação, constante a partir dos resultados das buscas, acompanham o desenvolvimento dos SRI e das expectativas dos usuários, agregando as evoluções lingüísticas correntes. Essas características trazem mobilidades justificando seu uso.

Verificou-se também que a evolução dos SRI tem mostrado uma preocupação crescente em contemplar requisitos de busca mais próximos ao modo de pesquisar do usuário. As interfaces se apresentam mais interativas, facilitando as operações. As linguagens de recuperação são mais objetivas e dinâmicas, cumprindo o objetivo de aproximar àquela utilizada pelo usuário do Sistema. Esta dinamicidade faz com que elas necessitem estar em constante avaliação, para que possam atender às exigências informacionais cada vez mais específicas e emergentes dos usuários. Também, para seu perfeito funcionamento, é preciso determinar parâmetros de validação que garantam sua consistência e aplicação como instrumento de navegação conceitual do usuário. A adoção de metodologias específicas de construção e validação, desenvolvidas com base em critérios pré-estabelecidos, vai justamente ao encontro dessa melhoria contínua.

Por conseguinte, acredita-se que os critérios devam ser de natureza lingüística, refletindo o uso da língua no contexto comunicativo, visto que se trata de uma linguagem utilizada para recuperar conteúdos cuja função é comunicar. A vantagem da aplicação de uma

metodologia de fundamentação lingüística sob a perspectiva do uso é a obtenção de uma ontologia atualizada, tanto no seu aspecto lexical como semântico, registrando variações de significante e significado e, assim, aumentar a eficiência da recuperação da informação.

O grande enfoque que vem sendo dado à utilização de ontologias nos SRI pode ser verificado em Congressos de Tecnologias da Informação, voltados para a busca de sistemas mais eficientes.

Por outro lado, por se tratar de temática recente, ainda em fase de aplicações tímidas e, em algumas situações, também controversas, a literatura carece de pesquisas que apresentem resultados efetivos, testados e documentados sobre a questão da validação dessas linguagens sob a ótica da Lingüística.

Ao abordar este tema, a presente pesquisa pretende contribuir para o projeto de construção de ontologias elaboradas para o sistema de informação do Observatório Base de Indicadores de Sustentabilidade Metropolitano de Curitiba (ORBIS_MC). Apresenta-se uma descrição detalhada desse projeto no Capítulo 3.

1.2 PROBLEMA DE PESQUISA

De maneira geral, é possível intuir e considerar que as informações que vão relatar a história dos povos, que vão expressar sentimentos e demonstrar conquistas, estas vão registrar o presente de uma sociedade para se tornar passado quase que no mesmo instante.

É fato, e amplamente encontrado na literatura corrente, que o atual crescimento científico, registrado com uma velocidade sem precedentes históricos, produziu, numa escala ainda maior, uma quantidade de dados e informações impossíveis de serem registrados ou recuperados na mesma velocidade de produção.

Esses dados e informações, disponíveis em suportes distintos, podem ser ordenados e classificados como: estruturados, semi-estruturados ou não estruturados. Nessa classificação, enquadram-se conteúdos disponíveis em sites de acesso livre, portais corporativos, sites governamentais, publicações diversas, documentos e bancos de dados, dentre outros. Essa diversidade vem então determinar a necessidade de desenvolvimento e disponibilização de formatos também diferenciados para a recuperação e possível utilização das informações. Nesse sentido, McGarry (1999, p. 11) afirma que “[...] A informação deve ser ordenada, estruturada ou contida de alguma forma, senão permanecerá amorfa e inutilizável. [...] A informação deve ser representada para nós de alguma forma, e transmitida por algum tipo de canal”.

Quanto à popularização dos conteúdos, hoje ela se faz por meio de mecanismos de acesso virtual, que facilitam e permitem a busca e consulta em catálogos, bases de dados ou outras bases que estejam disponibilizados em formato *on-line*, não exigindo a presença física do usuário. A busca de informações faz com que novas ferramentas sejam implementadas, aplicando a força das novas tecnologias que mudam conceitos, o modo de pensar, e a forma de realizar trabalhos e pesquisas.

Para McGarry (1999, p. 17), “[...] A linguagem é um sistema; uma rede de relações definidas”. A classificação da linguagem como um sistema relaciona os elementos entre si, onde tudo e todos têm, em algum momento, um ponto de relação e, para tal, precisam se comunicar com perfeição. Nesta prática de comunicação perfeita, há que se considerar as variações que o uso dos termos vão se agregando ao longo do tempo. Essas variações lingüísticas estarão representadas nos assuntos contidos nos documentos e registradas pela língua escrita.

No que concerne à recuperação de informação, essa tem uma ligação direta e indissolúvel com a representação dos assuntos, razão pela qual essas representações devem

expressar com fidelidade os pontos importantes, sendo quase que sínteses resumidas em palavras-chave ou descritores que vão possibilitar o acesso de forma rápida e objetiva. Esta representação tem caráter subjetivo, pois se espera que para cada leitura existam várias interpretações, diferentes análises, um novo olhar a cada leitor para cada nova leitura. Todo processo de leitura implica no posicionamento do leitor, fazendo com que a padronização da linguagem voltada à recuperação da informação se torne um trabalho direcionado a atender às necessidades dos usuários do sistema.

Dessa forma, a importância da linguagem de recuperação a ser utilizada está intimamente relacionada à análise documentária, específica para fins de indexação e recuperação, à sua forma de expressão e representação e à sua capacidade de evolução. A linguagem, elaborada e voltada para fins documentários, que encerra em si um sentido e uma estrutura própria, se traduz em si mesma e se volta a atender à expectativa do usuário, que tanto pode ser o produtor quanto o receptor das mensagens ou informações.

Para Cintra *et al.* (2002, p. 42), “[...] O vocabulário documentário tem por objetivo reunir unidades depuradas de tudo aquilo que possa obscurecer o sentido: ambigüidade de vocabulário ou de construção, sinonímia, pobreza informativa, redundância, etc”.

No planejamento das estratégias de buscas, são freqüentes os questionamentos que versam sobre a dificuldade de representação do conhecimento, sobre as linguagens a serem adotadas, sobre o resultado esperado quando da utilização da linguagem documentária ou natural, como melhor representar o conhecimento e como melhor representar a linguagem natural nas linguagens dos sistemas, estas ditas artificiais².

No mundo virtual, para se lidar com essa complexidade, observa-se uma crescente tendência no uso de linguagens classificadas como ontologias. A dificuldade apresenta-se em como elaborar vocabulários de representação do conhecimento, que empregam as ontologias

² Linguagem Documentária e Linguagem de Recuperação são tratadas como sinônimos. Linguagem de Sistemas é a linguagem de programação.

como linguagens, levando em consideração toda a gama de variações lexicais e semânticas do mundo real ou cotidiano. Sob o ponto de vista metodológico, a problemática se insere na definição correta de critérios lingüísticos a serem considerados na construção e validação das ontologias. Somente dessa forma acredita-se ser possível atingir os níveis adequados de revocação³ e precisão⁴ nos SRI.

Nesta pesquisa, o problema apresentado aplica-se ao Observatório Regional Base de Indicadores de Sustentabilidade Metropolitano de Curitiba (ORBIS_MC). O SRI, instalado no ORBIS_MC, tem por objetivo atender à necessidade de coleta de informações que levem a indicadores de desenvolvimento urbano e à qualidade de vida da Região Metropolitana de Curitiba e melhorias das ações públicas. As políticas públicas consideradas nesse contexto, e implementadas em conjunto com a iniciativa privada, pretendem o desenvolvimento uniforme da região visando o progresso futuro.

O ORBIS_MC tem por finalidade a coleta, o gerenciamento, a análise e o uso da informação de políticas públicas voltadas ao desenvolvimento econômico e social. Já o SRI buscará auxiliar, pelo uso da tecnologia, a melhoria de análise e emissão de relatórios, além de reunir diferentes tipos de informações em uma única base de conhecimento.

A opção por ontologias, como recurso a ser utilizado no SRI, tem por objetivo estabelecer o elo de comunicação entre o Sistema e a fonte das informações identificadas pelos usuários do sistema durante o processo de Engenharia do Conhecimento, que faz parte da metodologia de construção de ontologias. Com o auxílio das ontologias, acrescenta-se a tecnologia KMAI®⁵, que fará a coleta automática de dados ou informações com base no reconhecimento dos termos ou expressões criadas.

³ Revocação é a relação entre o número de documentos relevantes disponíveis no sistema e o número de documentos relevantes recuperados pelo usuário.

⁴ Precisão é a relação entre o número de documentos relevantes recuperados e o número total de documentos recuperados pelo sistema.

⁵ Uma descrição detalhada da tecnologia KMAI®, utilizada no SRI do ORBIS_MC, é fornecida no Capítulo 3.

A necessidade de elaboração de critérios lingüísticos para validação das ontologias adotadas está na importância da qualidade da coleta dos dados e informações, uma vez que os dados não coletados ou coletados de forma incorreta induzem ao erro dos índices obtidos estatisticamente. As buscas serão realizadas em fontes de mídia, por meio de observadores automáticos, onde as informações precisam ser reconhecidas pelo sistema, sendo que a linguagem a ser utilizada urge estar em perfeita sincronia com a linguagem natural apresentada nas fontes.

A elaboração de uma metodologia para a validação dessas ontologias e de seus relacionamentos semânticos buscará garantir a eficiência do Sistema no domínio estabelecido.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Propor uma metodologia para validação de ontologia por meio de critérios lingüísticos.

1.3.2 Objetivos específicos

- Definir critérios lingüísticos de validação da ontologia;
- Avaliar a qualidade das ontologias inseridas no sistema ORBIS_MC com base nos critérios lingüísticos de validação definidos.

CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA

2.1 TERMINOLOGIA

Ao longo da história, observou-se que todos os povos definiram formas de representar conteúdos, formas de se comunicar, de transmitir e receber mensagens. Essas tentativas de comunicação estão em registros históricos, desde os primórdios, onde os desenhos rupestres tinham um significado. Nesses desenhos, pode-se identificar o retrato da história dos povos, sua evolução e a representação dos mecanismos “tecnológicos” que foram sendo aperfeiçoados. Nesse aperfeiçoamento, inclui-se a padronização da nomenclatura, haja vista que os processos ou materiais são dotados de conceitos que devem ser retransmitidos.

Assim, a necessidade de se comunicar e se fazer entender têm levado à busca de padronizações que abarquem o conhecimento alvo de representação. Independente da língua utilizada ou da área do conhecimento, é indiscutível a necessidade de um conhecimento comum entre os pares.

Depois disso, para a análise da atividade de documentação, é preciso descrevê-la como um procedimento de tratamento e organização da informação, que sofre influência direta do meio onde está inserida. Na mesma condição também está a terminologia, que visa estabelecer uma possibilidade de comunicação entre produtor e leitor, gerando condições de elaborar significantes em cada área.

Segundo Krieger e Finatto (2004, p. 20-40),

[...] A Terminologia é uma disciplina que possui seu objeto primordial definido: o termo técnico-científico. [...] A Terminologia é um campo de estudos de caráter inter e transdisciplinar, o que a leva a convocar um conjunto de saberes para a apreensão do fenômeno terminológico, por excelência, o termo, cuja essência situa-se na representação lexical do conhecimento especializado e na sua divulgação. Para tanto, contribuem determinados conhecimentos exteriores e mesmo interiores aos estudos da linguagem (KRIEGER; FINATTO, 2004, p. 40).

Considerando essa perspectiva, é possível perceber que as comunicações científicas têm maior preocupação e cuidado sob este aspecto, pois se encontra com facilidade publicações que versam sobre as terminologias utilizadas em áreas determinadas. Por isso, Lyons (1987, p.33) diz que “[...] Toda disciplina dispõe de um vocabulário técnico próprio”. Por exemplo: Dicionários Terminológicos na área de Psicologia; Dicionário Terminológico de Inteligência Artificial; Dicionário Terminológico de Medicina; dentre tantos outros.

Para Krieger e Finatto (2004, p. 16), “[...] a terminologia é um termo polissêmico”. A polissemia se caracteriza pelo fato deste termo representar uma disciplina, o conjunto de termos representando o conhecimento especializado de uma determinada área e ainda a produção de glossários.

Além de estar relacionada à representação do conhecimento ou estar relacionada às comunicações científicas, a terminologia também está presente na corriqueira troca de palavras de grupos distintos. Considere-se, por exemplo, grupos de adolescentes de diversas épocas. As terminologias utilizadas para a comunicação entre eles, embora não estejam relacionadas para fins científicos, são práticas de comunicação.

Nesta pesquisa, considerou-se apenas a terminologia como uma forma de organização dos conceitos, na tentativa de facilitar a comunicação e a recuperação de informação.

Para los especialistas, la terminología es el reflejo formal de la organización conceptual de una especialidad, y un medio inevitable de expresión y de comunicación profesional. [...] Para los usuarios (directos e intermediarios) la terminología es un conjunto de unidades de comunicación, útiles y prácticas, cuyo valor se mide en función de criterios de economía, de precisión y de adecuación (CABRÉ, 1993, p. 37).

A autora demonstra a importância da terminologia para especialistas e usuários, que está voltada para a transmissão de informações especializadas que precisam ser consultadas, recuperadas, analisadas e utilizadas.

É certo que, para ser possível a interação entre especialistas e usuários, a linguagem terminológica deve estar adequada ao entendimento de ambos. Por essa razão, a terminologia pretende ser uma forma de padronização.

Para Krieger e Finatto (2004, p. 19), “[...] A tentativa de estabelecer uma padronização terminológica nas linguagens técnicas é própria da adoção de políticas lingüísticas articuladas sobre a crença de que o uso recorrente de um mesmo termo garante a univocidade da comunicação especializada”.

Assim, as terminologias têm a intenção de serem formas normalizadoras, objetivando um maior entendimento na comunicação e favorecendo a recuperação da informação. No entanto, nem sempre essa padronização é possível, uma vez que o uso terminológico por diferentes grupos de especialistas, de uma mesma área, pode trazer diferenças lexicais e semânticas.

2.2 LINGUAGENS DE ESPECIALIDADE

As terminologias são fundamentais para a perfeita representação e transmissão do conhecimento. Nesse sentido, as linguagens de especialidade são utilizadas por usuários de uma mesma área do conhecimento para agilizar a comunicação.

Para Santos e Sarmento (2005), as linguagens de especialidade são, genericamente, “[...] o conjunto de marcas lexicais, sintáticas, estilísticas e discursivas que tipificam o uso de um código lingüístico em qualquer ambiente de interação social”. Cabré (1993, p.133-134) utiliza definições de autores diversos para definir a linguagem de especialidade. Na interpretação da autora, são códigos de caráter lingüístico, ou uma simples variação da linguagem geral, ou subconjuntos para um entendimento amplo. Assim, com base nessas considerações diversas, a autora indica que: a) As linguagens tratam de conjuntos

especializados pela temática, pela experiência e universo de uso dos usuários; b) Elas se apresentam como um conjunto com características interrelacionadas e não como fenômenos isolados; c) Têm como predominante a função de comunicação, apesar de suas outras funções complementares (CABRÉ, 1993, p. 135).

Infere-se, assim, que as linguagens buscam essencialmente o intercâmbio de informações, em que pese, neste estudo, a necessidade de serem dinâmicas e contarem com o consenso dos pares. Contam com uma elaboração formal, porém não podem ser confundidas com linguagens documentárias, por conter características específicas.

Nesse sentido, Cabré (1993, p. 131) indica as razões que demonstram esta diferença:

- *Porque no se pueden considerar lenguajes <inventados>, en el sentido literal del término, sino que en buena medida participan de la lengua general;*
- *Porque no han sido contruidos tomando como punto de referencia el lenguaje, ya que propiamente forman parte de él;*
- *Porque pueden admitir nuevas unidades, aunque es deseable que se controle su entrada;*
- *Porque tienden (solo tienden) a disponer de una denominación para cada concepto, lo que les lleva a presentar un cierto grado de sinonimia no controlada;*
- *Porque disponen de todas las posibilidades sintácticas que les ofrece la lengua general, aunque suelen materializar solo una parte; y*
- *Porque cuentan con un repertorio de signos no necesariamente reducido.*

Como características, Cabré (1993, p. 131) sugere:

- *Cuentan com una conceptualización previa bastante controlada;*
- *No suelen admitir (em teoría) nuevas unidades si no están establecidas y conceptualizadas previamente;*
- *No tienen, en teoría, términos polisémicos; la polisemia del léxico común deviene homonimia en el léxico especializado;*
- *La forma escrita de los términos es prioritaria sobre su forma oral;*
- *Los términos especializados tienden a tener validez supranacional en muchos casos;*
- *No suele materializar ni la función emotiva ni la función poética del lenguaje, al menos en el grado en que ocurre en los textos del lenguaje general.*

As características demonstram que as linguagens, apesar de serem controladas, contam com a utilização de linguagens naturais e são destinadas a grupos bastante restritos na temática em que é construída. Essa restrição se manifesta nos SRI, que necessitam traduzir os conteúdos para uma linguagem amigável entre usuário e sistema.

2.3 UNIDADE TERMINOLÓGICA

2.3.1 Termo

Os termos encerram em si a propriedade de representar clara e explicitamente um conceito, favorecendo o entendimento ou a recuperação da informação.

Para Sager (1990, p. 57), “[...] *Terms are the linguistic representation of concepts*”⁶. Cabré (1993, p. 93) conclui que “[...] *los términos son unidades que relacionan el lenguaje y la realidad, que representan objetos de la realidad; con los términos, los individuos expresan e intercambian pensamientos y organizan la estructura de base de una disciplina*”.

Já Gouadec (1990, p.3) diz que “[...] *Un terme est une unité linguistique désignant un concept, un objet ou un processus. Le terme est l’unité de désignation d’éléments de l’univers perçu ou conçu. Il ne se confond que rarement avec le mot orthographique*”⁷.

No âmbito da Ciência da Informação, segundo Gomes (1990, p. 20), “[...] Entende-se por termo a palavra ou expressão selecionada para representar um dado significado”. Assim, pode-se inferir que os termos são utilizados para expressar conceitos específicos em áreas especializadas, sendo de amplo conhecimento e entendimento dos usuários.

Os termos também são empregados em situações específicas em que se pretende a comunicação técnica ou especializada acerca de determinado domínio. Na representação do conhecimento, os termos são recursos lingüísticos que trazem o conceito do que pretendem representar, considerando o contexto onde estão inseridos. Nesse sentido, Lara (2004, p. 92) diz que “[...] o termo é uma designação que corresponde a um conceito em uma linguagem de especialidade”. Cabré (1993, p. 169) acrescenta que “[...] *Los términos, que son las unidades*

⁶ Termos são a representação lingüística dos conceitos. (tradução nossa)

⁷ Um termo é uma unidade lingüística que designa um conceito, um objeto ou um processo. O termo é a unidade de designação de elementos do universo percebido ou concebido. Ele raramente se confunde com a palavra ortográfica. (tradução nossa)

de base de la terminología, designan los conceptos propios de cada disciplina especializada”. Para Felber (1984, p. 168), termo é “[...] *a linguistic symbol which is assigned to one or more concepts, which are defined from neighbouring concepts*⁸”. Como exemplo, o autor indica que o termo pode ser uma palavra ou um conjunto de palavras, uma letra ou um símbolo gráfico, uma abreviação, um acrônimo ou uma anotação, uma vez que os conceitos existem independentes dos termos.

Sendo os termos expressões de caráter técnico, científico ou tecnológico, é possível, por meio deles, a manifestação da realidade e do conhecimento, uma vez que se inserem em todos os textos e publicações especializadas. Assim, quando se avalia um texto, é possível identificar a diferença fundamental entre palavras e termos. Segundo Lara (2004, p. 92):

Uma palavra tem propriedades (como em um dicionário de língua), mas tem muitos significados, porquanto são elementos do léxico da língua. Um termo, ao contrário, é uma palavra contextualizada no discurso, tendo, conseqüentemente, um referente de interpretação.

Pode-se dizer então que os termos são manifestações lingüísticas e fazem parte das linguagens de especialidade, pois expressam um conhecimento de determinado domínio. No nível de sua estrutura, os termos podem se apresentar de forma simples, como “biodiversidade”, ou de forma complexa, como “óleo biocombustível”. Sua apresentação e utilização vão estar diretamente ligadas ao conceito contextualizado que representa.

2.3.2 Conceito

Segundo a norma ISO 704/2000 – Princípios e métodos da terminologia – os conceitos são construções mentais que vão categorizar os objetos e com os quais são representados nas variadas formas de comunicação, representado por (objeto → conceito → comunicação).

⁸ [...] um símbolo lingüístico que é designado a um ou mais conceitos que são definidos por conceitos contíguos. (tradução nossa)

Conforme a norma citada, para a terminologia, os conceitos são considerados representações mentais de objetos dentro de um contexto especializado. Não são, entretanto, confundidos com abstrações ou objetos imaginários, mas os elos de ligação entre o objeto e sua definição de forma abstrata.

A formação e utilização dos conceitos têm fundamental importância para organização do conhecimento, e é a partir dos conceitos que se torna possível a delimitação dos contextos, a classificação dos seres e dados, e a representação dos conteúdos.

De acordo com a sua representação, os conceitos podem ser classificados em:

- Individuais: quando descrevem um único objeto e são representados por uma linguagem especial que se torna marca ou símbolo. Ex.: internet; www.
- Gerais: quando descrevem um conjunto de dois ou mais objetos e a sua representação toma a forma de termos ou símbolo. Ex.: CD-Room; ®.

A literatura da área da Ciência da Informação define conceito como compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo lingüístico. São pressupostos indispensáveis na argumentação e nas comunicações verbais e constituem elementos necessários na construção de sistemas científicos (DAHLBERG, 1978). Cintra (2002) define a noção ou conceito como “unidade de pensamento constituído por propriedades comuns a uma classe de objetos” .

Na Lingüística, Cabré (1993) esclarece que conceitualização é um processo de seleção das características importantes que definem uma classe de objetos. A autora define conceito como o conjunto das características interrelacionadas que descrevem, essencial ou complementarmente, uma classe de objetos da realidade.

O conceito é o que identifica um objeto, ou seja, aquilo que encerra as características únicas que particularizam e remetem mentalmente a determinada forma ou definição. Quando se pretende dizer “água”, independente da língua em que se expresse, a noção a que se está

referindo é clara para todos ouvintes. O conceito de água tem características relevantes que a identificam e diferem de outros líquidos, sendo essas características amplamente conhecidas por qualquer pessoa.

Além de ter como objetivo identificar e diferenciar um objeto ou uma classe específica, os conceitos se diferenciam entre si pelas suas características ou propriedades. Assim, o conjunto de características determina uma classe e a ausência dela. Uma ou mais características ou propriedades permitem o estabelecimento das relações conceituais de forma hierarquizada. Para Cabré (1993, p. 198), as características se estabelecem por diferentes critérios:

- De acordo com a relevância, são: a) Essenciais: descrevem a essência e são imprescindíveis para garantir sua definição; b) Complementares: não essenciais, adicionam elementos à descrição.
- De acordo com a relação com o objeto: a) Intrínsecas: representam a descrição relacionada a uma classe, tais como: forma, cor, tipo de ação, etc.; b) Extrínsecas: externas à sua definição, tais como: função, origem, localização, etc.

As características de um conceito vão determinar a classificação dos termos, estabelecendo as relações conceituais que se dispõem hierarquicamente, partindo de conceitos com características mais abrangentes para as características mais específicas.

2.4 RELAÇÃO ENTRE SIGNIFICANTE E SIGNIFICADO

Quando se pretende a comunicação, é certo que os discursos se constroem com palavras, que, para Saussure (2003, p. 79-81), são definidas como signos lingüísticos. Os signos são formados pelo conceito e pelo som, que classifica como significado e significante. Sendo intrinsecamente relacionados, já que um depende do outro, eles detêm a

funcionalidade de exprimir as características que vão diferenciar as palavras. Esta relação, entretanto, se dá de forma bastante arbitrária, não determinando que o som tenha sempre o mesmo significado, em que pese a diferença das línguas praticadas. Assim, não existe uma obrigatoriedade para que um som remeta a um conceito.

Saussure (2003, p. 81) diz que “[...] O laço que une o significante ao significado é arbitrário ou então, visto que entendemos por signo o total resultante da associação de um significante com um significado, podemos dizer mais simplesmente: o signo lingüístico é arbitrário”.

Para Lyons (1987, p. 103), tradicionalmente os significados “[...] são idéias ou conceitos que se podem transferir da mente do falante para a do ouvinte por encarnar-se, por assim dizer, nas formas de uma ou outra língua”.

Encontra-se em Cabré (1993, p. 213) a indicação de que as relações entre significante e significado ocorrem de forma múltipla. Desta forma, uma unidade pode ter vários significados ou um conceito pode ser denominado de várias formas.

2.4.1 Homonímia

Conforme define o Dicionário de Lingüística (1986, p. 326), a homonímia é “[...] a identidade fônica (homofonia) ou a identidade gráfica (homografia) de dois morfemas que não têm o mesmo sentido, de um modo geral.” Segundo Câmara Júnior (1996, p. 139), homonímia se define como: “Propriedade de duas ou mais formas, inteiramente distintas pela significação ou função, terem a mesma estrutura fonológica: os mesmos fonemas, dispostos na mesma ordem e subordinados ao mesmo tipo de acentuação”. Para Cintra (2002, p. 71), a homonímia “[...] consiste em uma mesma forma significante remeter a duas realidades vocabulares diversas, sejam unidades com identidade fônica (homofonia) ou identidade

gráfica (homografia)”. A autora relata que esta relação entre significante e significado pode gerar ambigüidade. Para Lyons (1987, p.110), “[...] homônimos são palavras diferentes (i.e.lexemas) com uma forma igual”. Conceito semelhante encontra-se na Gramática Portuguesa que diz: “[...] É a relação entre duas ou mais palavras que, apesar de possuírem significados diferentes, possuem a mesma estrutura fonológica”.

Ainda, segundo a gramática, as palavras homônimas se classificam em perfeitas e imperfeitas.

- Homógrafas heterofônicas (homógrafas): são as palavras iguais na escrita e diferentes na pronúncia. (Ex.: conserto: substantivo; conserto: verbo consertar);
- Homófonas heterográficas (homófonas): são as palavras iguais na pronúncia e diferentes na escrita. (Ex.: cela: substantivo; sela: verbo);
- Homófonas homográficas (homônimos perfeitos): são as palavras iguais na pronúncia e na escrita. (Ex.: cura: verbo; cura: substantivo).

2.4.2 Polissemia

A polissemia indica uma diversidade de significados. A definição do Dicionário de Lingüística diz que: “Chama-se polissemia à propriedade do signo lingüístico que possui vários sentidos”. Segundo a Gramática Portuguesa, o termo polissemia “[...] É a propriedade que uma mesma palavra tem de apresentar vários significados. Ex.: posto de gasolina; posto na empresa”. De acordo com o Glossário de Gramática e Lingüística, polissemia é “[...] Fenômeno lingüístico pelos quais vários significados inter-relacionados, surgidos a partir de um primitivo ou original, associam-se ao mesmo significante”.

Essas transformações lingüísticas dizem respeito ou estão inseridas em um contexto específico, como define Câmara Júnior (1996, p. 194): a polissemia é a “[...] Propriedade da

significação lingüística de abarcar toda uma gama de significações, que se definem e precisam dentro de um contexto”.

2.4.3 Sinonímia

De acordo com a Gramática da Língua Portuguesa, uma relação sinonímica é aquela “[...] que se estabelece entre duas palavras ou mais que apresentam significados iguais ou semelhantes (Ex: distante; afastado)”. Para Lyons (1987, p.111): “[...] Se a sinonímia for definida como identidade de significado, poderemos dizer que os lexemas são completamente sinônimos (em uma certa faixa de contextos) se, e somente se, tiverem o mesmo significado descritivo, expressivo e social (na faixa de contextos em questão)”. Cintra (2002, p. 75) afirma que sinonímia “[...] pode ser definida como identidade de significação entre elementos lexicais, porém, a existência de sinonímia absoluta é controversa”.

O entendimento de sinonímia se faz em dois aspectos diferentes, segundo Dubois *et al.* (1986, p. 555): quando os termos têm a possibilidade de se substituírem um ao outro ou quando dois termos são intercambiáveis em todos os contextos. Os autores concluem que a sinonímia é “[...] a tendência das unidades do léxico de terem o mesmo significado e de serem substituíveis umas pelas outras”. No mesmo sentido, está o proposto na ISO 704:2000 (p. 25) que define:

*Synonymy is the relation differing designations that designate the same concept, i.e., having the same intension, in a given language. Designations in a synonymous relation are called synonyms. Given the same level of language, synonyms are interchangeable. If two or more terms are attributed to concepts whose intensions are almost identical, they are called quase-synonyms and are interchangeable only in some contexts.*⁹

⁹ Sinonímia é a relação entre designações diferentes que designa o mesmo conceito, i.e., tendo a mesma intensão em uma determinada língua. Designações em uma relação sinonímica são chamadas sinônimas. Dado o mesmo nível de linguagem, sinônimos são intercambiáveis. Se dois ou mais termos são atribuídos a conceitos cujas intensões são quase idênticas, são chamados quase-sinônimos e são intercambiáveis só em alguns contextos”. (tradução nossa)

De acordo com o Vocabulário de Terminologia, incluído na ISO Recommendation ISO/R 1087-1969 (p. 336), “*synonymous (several terms): Having the same meaning [...] exactly or nearly, but different external forms [...]*”¹⁰.

Para Faulstich (2005, p. 42), a sinonímia pode ser considerada sob dois aspectos: quando dois termos podem ser substituídos mutuamente ou quando são intercambiáveis em todos os contextos.

Na sinonímia, insere-se a hiponímia, que, quando se apresenta como uma relação recíproca e simétrica, é classificada como sinônima. Assim, como exemplifica Faulstich (2005, p. 43), “[...] Em determinado contexto, onde bate-bola é igual a jogo e jogo é igual a bate-bola a relação entre os termos é recíproca, logo sinonímica.”

Entretanto, normalmente, a hiponímia é assimétrica. Faulstich (2005, p. 42-43) diz que “[...] A hiponímia deve ser entendida como relação de inclusão de significados das unidades em questão, assim é que o subconjunto (assassino, matador, pistoleiro, carrasco, bandido, celerado) está incluso no conjunto criminoso”. Continua a autora: a hiponímia, propriamente dita, se define “[...] por uma relação de implicação unilateral, assim é que se um objeto é esverdeado pode-se entender que esse objeto seja verde, mas se o objeto é verde não se diz que ele é esverdeado.”

Essa relação é que determina que a hiponímia é assimétrica, como também identifica Cintra (2002, p. 77), em que a hiponímia está caracterizada linguisticamente dentro das relações de sentido, considerando a significação. Assim, a hiponímia se refere à noção de inclusão, que reúne unidades em uma mesma classe.

Para Lyons *apud* Cintra (2002, p. 78), “a hiponímia expressa a relação existente entre um lexema mais específico ou subordinado, e um lexema mais geral ou superordenado, [...]”

¹⁰ Sinônimos (vários termos): têm o mesmo significado [...] exatamente ou próximo, mas com diferentes formas externas [...]. (tradução nossa)

Verifica-se, pelas definições, que a hiponímia vai determinar a lógica de pertinência das classes ou gênero dentro de um contexto.

2.5 RELAÇÕES ENTRE CONCEITOS

Os conceitos, obrigatoriamente, se relacionam quando se identificam características semelhantes entre eles. Para Dahlberg (1978, p. 104), “[...] Sempre que diferentes conceitos possuem características idênticas, deve-se admitir que entre eles existem relações”.

A partir das características básicas dos conceitos, que podem ser essenciais quando estão relacionadas às finalidades e aplicação ou acidentais quando dizem respeito às condições externas, verifica-se a sua importância quando se pretende a representação do conhecimento. A representação se vale dessas características para elaborar a sistematização dos conteúdos.

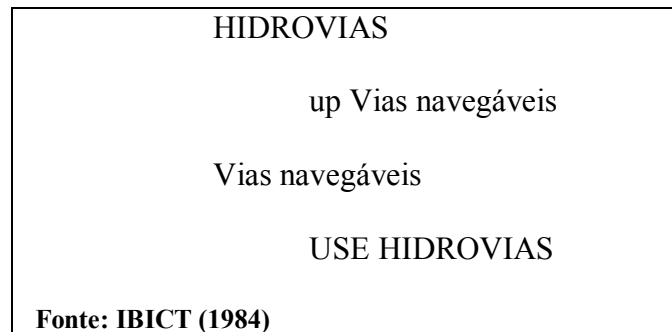
Esta representação considera o contexto do domínio-alvo, no qual todos os termos com características pertinentes estabelecem uma relação. Cabré (1993, p. 201), considerando a caracterização dos termos, indica como as relações ocorrem:

Considerados em conjunto, los términos no son de ninguna forma unidades aisladas que puedan darse fuera de un contexto específico, sino elementos que forman parte de un sistema lingüístico especializado y que se materializan en un contexto concreto que corresponde a un campo determinado de especialización. [...] un término se relaciona con todos los demás términos que forman parte de un mismo campo especializado, con los que constituye un campo conceptual o campo nocional.

Segundo Gomes (1990, p. 39), quando instrue para a elaboração de tesouros, o relacionamento entre conceitos é uma estruturação. Assim, “[...] sempre que houver características comuns (presentes na definição), existe um relacionamento direto entre os conceitos”.

2.5.1 Relações de equivalência

As relações de equivalência entre “[...] dois ou mais termos são considerados, para fins de indexação, como designando o mesmo conceito” (IBICT, 1984 p.22). Nesse contexto, há entre os termos uma reciprocidade; por exemplo:



2.5.2 Relações lógicas

As relações lógicas se estabelecem pela semelhança entre os conceitos. Como semelhança, são consideradas as características comuns aos dois conceitos alvo do relacionamento. Para elaboração de tesauros, segundo Gomes (1990, p. 41), o relacionameto lógico “[...] é um relacionamento de similaridade porque, quando se comparam dois conceitos entre si, verifica-se se possuem algumas características em comum”.

Cabré (1993, p. 202) considera a comparação e a combinação dos conceitos quando elabora sobre relacionamentos lógicos. Segundo a autora,

Las relaciones lógicas que se establecen entre los conceptos se basan en la semejanza que estos mantienen entre sí, por el hecho de poseer una o más características comunes: esta semejanza se manifiesta tanto cuando se comparan distintos conceptos (la comparación) como cuando se combinan entre sí (la combinación) (CABRÉ, 1993, p. 202).

Quanto às características, Dahlberg (1978, p. 104) estabelece que as relações lógicas podem acontecer entre dois conceitos, chamados A e B, nas seguintes situações:

- Identidade: onde as características são as mesmas entre os dois conceitos;

- Implicação: conceito de A está contido no conceito de B;
- Intersecção: algum elemento coincide nos dois conceitos;
- Disjunção: não apresentam nenhuma característica em comum;
- Negação: conceito de A tem característica que encontra a negação em B.

Essas classificações permitem a elaboração e o entendimento dos relacionamentos lógicos na representação do conhecimento e a organização dos conteúdos. Assim, segundo Dahlberg (1978), considerando as características dos conceitos, as relações podem ser hierárquicas (genéricas e partitivas), de oposição e funcionais.

2.5.2.1 Relação hierárquica

As relações hierárquicas são aquelas que se definem nas noções subordinadas em um ou vários níveis (ISO 1087). São compostas pelas relações genérico/específico e as partitivas.

2.5.2.1a Relação genérica

Caracterizadas como de gênero e espécie, as relações genéricas se estabelecem quando dois conceitos apresentam características semelhantes, porém um dos conceitos tem uma característica a mais que o outro. Quando um conceito é mais abrangente ou mais genérico e “absorve” ou contém o outro, fica estabelecida uma relação hierárquica. Pode-se dizer também que um termo na hierarquia sempre é mais completo que outro.

Para Gomes (1990, p. 41), a ordenação dos termos, a partir da relação hierárquica, determina uma verticalização, em que os termos superordenados (com todas as características) se ligam a termos subordinados (com pelo menos uma característica a mais).

2.5.2.1b Relação partitiva

As relações partitivas se estabelecem entre um objeto, considerado todo, e suas partes subordinadas ou que o constituem. Essas relações existem quando o conceito superordenado representa um total, enquanto os conceitos relacionados ou inferiores representam as partes deste total (ISO 704/2000). Segundo Sager (1990, p. 32), as relações partitivas servem para indicar a conexão entre conceitos, consistindo num todo e suas partes e podem ser representadas por árvores. Para Cabré (1993), são relações onde o todo e as suas partes se relacionam. Também chamadas de relações “parte-todo”, elas se baseiam na contigüidade dos objetos no espaço, ou seja, simultaneamente.

Conforme a norma de elaboração de tesauro, Gomes (1990, p. 44) apresenta quatro tipos de classes que representam o relacionamento de todo/parte: sistemas e órgãos do corpo; localidades geográficas; disciplinas ou áreas do discurso; e estruturas sociais hierarquizadas.

Para Felber e Picht (1984), as relações partitivas podem ser estabelecidas em séries horizontais ou verticais e são verificadas apenas na ordenação das partes de um conceito ou de um objeto individual.

As relações partitivas ocorrem quando se apresenta um todo e suas partes, por exemplo: quando se fala em avião, como pertinentes, têm-se as partes (asa, hélice, fuselagem, turbina, trem de pouso) ou quando as partes relacionadas fazem parte de um todo, como: pneu, aro, pedal, guidão e selim fazem parte da bicicleta.

2.5.3 Relação associativa

As relações associativas são de difícil definição. É possível encontrar na literatura definições que as caracterizem como aquelas relações que não podem ser classificadas como hierárquicas ou como sinonímicas.

Para Gomes (1984, p.31), esses relacionamentos ocorrem entre termos que não são equivalentes nem formam uma hierarquia, em que um termo seja subordinado a outro; contudo, são tão associados mentalmente que se deve tornar esta ligação explícita. Definição semelhante pode ser encontrada na ISO 704/2000 (p.12) que diz: “[...] *Associative relations are non-hierarchical. An associative relation exists when a thematic connection can be established between concepts by virtue of experience*”¹¹. Já a ISO 2788/1974 propõe que “[...] *The associative relation is usually employed to cover the other relations between concepts that are related but are neither consistently hierarchical nor equivalent (e.g. similarity, antonymity)*”¹².

Encontra-se em Cintra (2002, p.63) uma breve indicação da dificuldade na definição das relações associativas, uma vez que, em algum momento, todos os termos, palavras ou conceitos estabelecem um relacionamento considerando o universo ou o domínio onde estão inseridos. Essa dificuldade também é alvo de pesquisa e análise de Marroni (2006, p. 1), que considera a inexistência de uma delimitação das relações como fator comprometedor para a representação e recuperação dos conteúdos. Nesse sentido, segundo o autor, busca “identificar os parâmetros capazes de explicitar mais adequadamente essas relações associativas”.

Nas relações associativas, encontram-se, entre outras, as relações de oposição que Dahlberg (1978) cita como uma das relações possíveis a partir das características dos conceitos e que expressam propriedades. Elas podem ser: a) de contradição: presente – ausente; b) de contrariedade: branco – preto.

Outro tipo de relação associativa são as chamadas relações funcionais. Essas relações se dão a partir de conceitos que expressam processos. Para Dahlberg (1978, p. 105), elas têm

¹¹ Relações associativas não são hierárquicas. Existem quando uma conexão temática pode ser estabelecida entre conceitos em virtude da experiência. (tradução nossa).

¹² A relação associativa normalmente é empregada para tratar de outras relações entre conceitos que são relacionados mas não são nem consistentemente hierárquicos nem equivalentes. (tradução nossa)

por base as valências semânticas dos verbos. O autor entende por valência semântica “a soma dos lugares a serem preenchidos de acordo com a ligação deste conceito com outro”.

2.6 VARIANTES TERMINOLÓGICAS

Para proceder à análise de terminologias, Faultstich (1996) expõe que “[...] Em terminologia, os fenômenos variáveis ocorrem no sistema interno da língua na qual estão redigidos os textos de especialidade”. Essa definição aponta que, para se proceder a uma avaliação terminológica, deve-se considerar o texto na sua totalidade, a contextualização dos termos, para então elaborar sobre as variações, o quê e como as terminologias variam.

Indicação semelhante faz Sager (1990, p. 213), quando sugere que “[...] *Observation of usage permits the identification and categorisation of linguistic variants of terms by text types*¹³”. Para o autor, é importante observar que a variação de categoria dos termos tem relação direta com a densidade dos termos num determinado texto, considerando seu contexto.

Uma classificação das variantes é proposta por Faultstich (1996) que as divide em dois grandes grupos: 1) Variantes Lingüísticas: o fenômeno lingüístico determina o processo de variação; 2) Variantes de Registro: a variação ocorre nos planos horizontal, vertical e temporal da língua.

Para a classificação das variantes lingüísticas, a autora considera: a) A interpretação semântica é a base para análise do termo; b) As unidades terminológicas complexas são analisadas sob o ponto de vista funcional; c) Os subsistemas da língua portuguesa constituem fundo lingüístico de análise; d) Os usos escrito e oral dos termos são levados em conta.

¹³ A observação do uso permite a identificação e categorização de variants lingüísticas de termos por tipo de texto.

Para a classificação das variantes de registro, a autora considera: a) Os termos são recolhidos no discurso real da linguagem de especialidade; b) Os termos pertencem à variedade socioprofissional; c) Os termos são recolhidos de textos, de procedência diversificada, mas que tratam do mesmo assunto; d) Os termos são recolhidos de discursos com maior ou menor grau de formalização, que tratam do mesmo assunto; f) Os termos são recolhidos de textos redigidos em épocas diferentes, que tratam do mesmo assunto; g) Os usos escrito e oral são levados em conta.

A partir desses critérios e seguindo os princípios de classificação propostos, as variantes se subdividem e apresentam a seguinte classificação:

VARIANTES LINGÜÍSTICAS

- Variante terminológica morfossintática: na constituição do termo ocorre a alternância morfológica e sintática sem que o conceito se altere;
- Variante terminológica lexical: a forma lexical se modifica mas não altera o conceito do termo;
- Variante terminológica gráfica: a diversidade gráfica está ligada às convenções da língua.

VARIANTES DE REGISTRO

- Variante terminológica geográfica: ocorre no plano horizontal de diferentes regiões em que se fala a mesma língua;
- Variante terminológica de discurso: decorre da sintonia entre elaborador e usuários de textos mais formais ou menos formais;
- Variante terminológica temporal: concretiza-se como mudança num discurso escrito em épocas diferentes sobre um mesmo assunto.

No processo de comunicação, diversos fatores contribuem para as variações terminológicas, determinando que um mesmo termo tenha conceitos diferentes ou diversos termos tenham o mesmo conceito, considerando-se a importância da análise do contexto.

2.7 LINGUAGENS

A linguagem é a forma de comunicação entre os grupos e destes com o conjunto da sociedade. Nessa perspectiva de comunicação, as linguagens são também a forma de interação entre os modernos recursos tecnológicos e o usuário, no binômio amplamente difundido homem-máquina.

2.7.1 Linguagem natural

Ao longo da última década, o volume crescente da produção de conteúdos vem gerando a necessidade de possibilitar ao usuário autonomia suficiente para recuperar informações. A diversificação dos SRI disponíveis espelha esta nova realidade.

Na tentativa de estabelecer interface amigável entre sistemas e usuários, as pesquisas têm considerado a linguagem natural como base de comunicação e espelho da cultura de determinada sociedade. Nesse sentido, Dahlberg (1978, p. 101) teoriza dizendo que “...a linguagem constitui a capacidade do homem designar os objetos que o circundam assim como de comunicar-se com seus semelhantes” (grifo do autor).

Para Lancaster (1993, p. 200), a linguagem natural (LN) pode ser definida como sinônimo de ‘discurso comum’, ou “[...] a linguagem utilizada habitualmente na escrita e na fala, e que é o contrário de vocabulário controlado”.

Sendo uma expressão de um grupo e sua cultura, Cintra (2002, p. 15) indica que “[...] Na LN, o elemento de troca é o signo lingüístico que associa um significante (imagem acústica) a um significado (conceito)”.

Nos processos de indexação, a LN serve de base para a elaboração das linguagens documentárias (LD), que, padronizadas, pretendem uma recuperação satisfatória para o usuário. Pesquisas recentes têm procurado demonstrar como a aplicação da linguagem natural vem sendo utilizada, enquanto recurso de busca, e quais os resultados nos sistemas disponíveis.

2.7.2 Linguagem documentária

Segundo Cintra (2002, p. 33), a utilização das linguagens documentárias (LD), como recurso para recuperação da informação, vem das décadas de 1950 e 1960, com o crescimento do conhecimento científico e tecnológico.

Guinchart e Menou (1994, p. 133) definem linguagem documentária como “[...] uma linguagem convencional utilizada por uma unidade de informação para descrever o conteúdo dos documentos, com o objetivo de armazená-los e recuperar as informações que eles contém”. Cintra (2002, p. 33) define como “[...] linguagens construídas para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a ‘traduzir’ os conteúdos dos documentos”. Para Gardin (1968) *apud* Cintra (2002, p. 35), uma LD é “[...] um conjunto de termos, providos ou não de regras sintáticas, utilizadas para representar conteúdos de documentos técnico-científicos com fins de classificação ou busca retrospectiva de informações”.

Essas linguagens são “construídas” ou elaboradas com a finalidade de garantir o entendimento dos conteúdos a partir de uma normalização da LN e são considerados

“instrumentos intermediários” (CINTRA, 2002). A autora busca em Gardin *et al.* (1968) os elementos básicos que integram uma LD:

- Um léxico, identificado como uma lista de elementos descritores, devidamente filtrados e depurados;
- Uma rede paradigmática para traduzir certas relações essenciais e, geralmente estável, entre descritores. Essa rede lógico-semântica corresponde à organização dos descritores numa forma que, *lato sensu*, poder-se-ia chamar de classificação;
- Uma rede sintagmática destinada a expressar as relações contingentes entre os descritores, relações que são válidas no contexto particular onde aparecem. A construção de “sintagmas” é feita por meio de regras sintáticas destinadas a coordenar os termos que dão conta do tema.

As linguagens utilizadas no SRI encerram o objetivo de permitir a indexação e recuperação dos conteúdos. Assim, com o controle das terminologias e a padronização dos termos, torna-se possível encontrar respostas para as questões formuladas pelo usuário no momento da busca.

Por se tratarem de linguagens elaboradas, elas se valem da técnica de coordenação dos termos a partir dos conceitos. Essa coordenação pode ser feita em dois momentos distintos: 1) quando da indexação, feita pelo técnico especialista e classificadas como linguagens pré-coordenadas; 2) ou quando da recuperação, feita pelo usuário no momento da busca e classificadas como pós-coordenadas.

- Linguagens pré-coordenadas: são utilizadas quando a coordenação dos termos é feita no momento da indexação, na qual são considerados vários tópicos do assunto a ser representado e a importância destes para o usuário. A indexação obedece a critérios subjetivos, uma vez que o técnico é que determina o quê e como será representado o

conteúdo. Ex.: Classificação Decimal de Dewey (CDD); Classificação Decimal Universal (CDU).

- Linguagens pós-coordenadas: são utilizadas quando a coordenação dos termos é feita pelo usuário no momento da busca, de forma não ordenada e de acordo com as necessidades do usuário. Para proceder a busca, todos os termos são considerados, não representando maior ou menor importância entre eles. Essas linguagens são comumente utilizadas nos sistemas de busca automatizados, que utilizam elementos da lógica booleana, tais como *and* e *or* para realizar a ligação dos termos. Ex.: Tesouros.

As linguagens utilizadas nos SRI são fundamentais para o seu desenvolvimento e para a interação com o usuário. Considerando que os sistemas precisam atender às expectativas dos usuários, o estudo exaustivo das linguagens a serem utilizadas deve considerar os objetivos do sistema, a área do conhecimento que pretende representar e as necessidades dos possíveis usuários.

2.8 ONTOLOGIAS

A palavra ontologia deriva do grego *onto* (ser) e *logia* (discurso escrito ou falado). Tradicionalmente, quando se pretende definir ontologia, busca-se na Filosofia que diz “o estudo do ser enquanto ser” ou a “teoria do ser”, e estuda as coisas como elas existem ou como são. Pode-se dizer que as ontologias favorecem ou permitem o entendimento de como as ligações se estabelecem no mundo.

As correntes filosóficas vêm se dedicando ao estudo da ontologia como disciplina desde o século XVIII. Conforme relata Feitosa (2005, p. 210), três correntes filosóficas importantes surgiram a partir do século XIX. Segundo Husserl *apud* Feitosa (2005), a ontologia é a ciência das essências e pode ser de dois tipos: 1) formal: que é o fundamento de

todas as ciências e interessa-se pelas essências e suas relações com outras essências; 2) material: conjunto de ontologias setoriais que se preocupa com fatos. Para Hartmann *apud* Feitosa (2005), a ontologia presta-se ao reconhecimento daquilo que é metafisicamente insolúvel; essa corrente examinava os diversos momentos do ser, relacionando-os à existência ou à essência; à realidade ou à idealidade; à possibilidade, realidade, causalidade ou necessidade. De acordo com Heidegger *apud* Feitosa (2005), a ontologia é uma metafísica da existência; é baseada na fenomenologia.

Na literatura, é possível encontrar outras definições. Para a Filosofia, é uma descrição sistemática da Existência. Na Inteligência Artificial (IA), é uma especificação formal explícita de como representar os objetos, conceitos e outras entidades que supostamente existem em alguma área de interesse e os relacionamentos que existem entre eles. Para os sistemas de IA, o que “existe” é o que pode ser representado. Então, a ontologia pode ser um programa que define um conjunto de termos a serem representados.

Sua adaptação para a Ciência da Computação adiciona ao termo taxonomia às especificações das relações entre as entidades mais um conjunto de regras automáticas de inferência e ações associadas (PARREIRAS, 2004).

Outras descrições são atribuídas ao termo ontologia, conforme a área de atuação. Na área da Ciência da Informação,

[...] as ontologías establecen una terminología conjunta entre miembros de una comunidad de interés. Estos miembros pueden ser humanos os agentes automáticos. Una ontología es una especificación de una conceptualización. [...] Ontología es una descripción formal de conceptos en el dominio de un discurso (CAMARA, 2002).

Na Gestão do Conhecimento, Guillén, Moreno e Badaya (2005) afirmam que:

[...] as ontologías se encasillan dentro de las LISTAS RELACIONADAS, como tipo de sistema de ORGANIZACIÓN DEL CONOCIMIENTO. Una ONTOLOGÍA define el vocabulario de un área mediante un conjunto de términos básicos y relaciones entre dichos conceptos, así como las reglas que combinan términos y relaciones. [...] La GESTIÓN DEL CONOCIMIENTO está desarrollando ontologías como modelos de conceptos específicos. Pueden representar relaciones complejas entre los objetos, e incluyen las reglas y los axiomas que faltan en los tesauros.

Relacionado à área tecnológica, o Núcleo de Estudos em Tecnologias para Informação e Conhecimento define ontologia como sendo:

Especificações das relações entre as entidades mais um conjunto de regras automáticas de inferência e ações associadas. É uma descrição formal dos conceitos e relacionamentos que existem dentro de um domínio. Isso significa que uma ontologia se relaciona com um vocabulário específico, uma linguagem específica e a conceitualização de determinado domínio (NETIC, 2003).

Na Inteligência Artificial (IA), é possível identificar em Gruber (1993) que as ontologias se apresentam como linguagens ou formas de representação do conhecimento que possibilitam a comunicação entre os SRI e os usuários, além de favorecer o entendimento dos conteúdos pelo compartilhamento do conhecimento, representado nos termos ou expressões construídas com base nos conceitos. É uma forma de representar e organizar, considerando os princípios semânticos.

Voltados para a estruturação e representação do conhecimento, Swartout *et al.* (1996) se referem à ontologia como um conjunto de termos estruturados que descrevem algum domínio e são a estrutura para uma base de conhecimento.

Gruber (1993) elaborou o conceito que melhor define ontologias: “*An ontology is na explicit specification of a conceptualization*”¹⁴. O autor, a partir dessa definição, sugere a condição das ontologias que permitem a reutilização do conhecimento e indica a sua apropriação pela Inteligência Artificial.

*For AI systems, what “exists” is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, an ontology is the statement of a logical theory*¹⁵.

¹⁴ Ontologia é a especificação explícita de uma conceitualização. (tradução nossa)

¹⁵ Para sistemas de IA, o que “existe” é o que pode ser representado. Quando o conhecimento de um domínio é representado num formalismo declarado, o conjunto de objetos que pode ser representado é chamado de universo de discurso. Este conjunto de objetos, e seus relacionamentos possíveis de serem descritos, são refletidos no vocabulário de representação com o qual um sistema baseado em conhecimento representa o conhecimento.

O termo ontologia é associado diretamente à descrição de um conhecimento específico, apresentando-se como forma de representação do domínio do conhecimento através de um vocabulário construído para esta finalidade. Esse vocabulário vai abranger os conceitos e as relações que podem existir dentro do domínio.

Quando se pretende a representação do conhecimento e a conseqüente construção de ontologias, alguns critérios devem ser considerados, a partir da conceitualização, que vão definir a estrutura e a sua elaboração, voltadas ao compartilhamento do conhecimento e a recuperação (GRUBER, 1993):

- Clareza: a ontologia deve expressar com objetividade e em linguagem natural o significado dos termos definidos.
- Coerência: deve validar inferências que são coerentes com as definições e esta coerência deve estar presente também nos conceitos expressos em linguagem natural.
- Extensibilidade: a ontologia deve estar voltada para o compartilhamento do uso, permitindo a sua atualização e inclusão de novos termos.
- Codificação parcial: a conceitualização deve ser especificada no nível de conhecimento sem depender de um nível simbólico.
- Comprometimento ontológico mínimo: deve permitir as atividades de partilhamento do conhecimento em um contexto menor possível, possibilitando a modelagem do contexto com liberdade e agilidade.

Nessa perspectiva, Gómez-Pérez (2000, p. 34) acrescenta outros critérios aos expostos por Gruber, aos quais chama de princípios, que também devem ser considerados na construção de ontologias:

Assim, no contexto de IA, nós podemos descrever a ontologia de um sistema definindo um conjunto de termos de representação. Em tal ontologia, as definições associam os nomes de entidades do universo de discurso (por exemplo, classes, relações, funções, ou outros objetos) com o texto legível pelo homem descrevendo o que os nomes significam, e axiomas formais que determinam a interpretação e uso correto destes termos. Formalmente, uma ontologia é a declaração de uma teoria lógica. (tradução nossa)

- Princípio da distinção ontológica: as classes de uma ontologia devem ser desmembradas;
- Diversificação de hierarquias: aumentar a capacidade fornecida por múltiplos mecanismos herdados, sendo possível entrar com novos conceitos e também herdar propriedades a partir de diferentes pontos de vista;
- Modularidade: para minimizar a união entre os módulos;
- Minimizar a distância semântica entre conceitos similares: conceitos similares são agrupados e representados como subclasses de uma classe e devem ser definidos usando as mesmas primitivas, enquanto conceitos, que são menos similares, são representados mais distantes na hierarquia;
- Padronizar nomes sempre que possível.

As ontologias, por meio de suas características, pretendem abranger todo o domínio-alvo de sua construção, representando o conhecimento através de termos de entendimento comum e suas respectivas relações, assim como contribuindo para a facilidade de recuperação das informações ali representadas.

Nesse sentido, Gómez-Pérez (2000, p. 35) diz que as ontologias fornecem um vocabulário comum de uma área e definem – com diferentes níveis de formalismo – o significado dos termos e as relações entre eles. Para a autora, os conhecimentos representados pelas ontologias são formalizados utilizando cinco tipos de componentes:

- 1) Classes: os conceitos são usados em sentido amplo, podendo ser abstratos ou concretos, elementares ou compostos, reais ou fictícios e podem ser algo sobre o que é dito e também pode ser a descrição de uma tarefa, função, ação, estratégia, argumentação;
- 2) Relações: representam um tipo de interação entre os conceitos de um domínio; formalmente definidos como qualquer subconjunto de um produto de ' n ' conjuntos;

- 3) Funções: é um caso especial de relações no qual o n -ésimo elemento da relação é único para os ' $n-1$ ' elementos precedentes;
- 4) Axiomas: são usados para modelar as sentenças que são sempre verdadeiras;
- 5) Instâncias: usadas para representar elementos.

No processo de construção de ontologias, torna-se imprescindível considerar o contexto onde serão inseridas e os objetivos do sistema, representando, assim, a razão para a qual serão elaboradas e empregadas. Uma proposta de metodologia única para esta construção deve considerar as seguintes dimensões propostas por Uschold (1996, p. 3):

- Formalidade: o grau de formalidade da representação do vocabulário e da especificação do significado pode ser: a) altamente informais: expressadas livremente em linguagem natural; b) informal estruturada (semi-informal): expressada em linguagem natural, de forma restrita e estruturada, reduzindo a ambigüidade; c) semiformal: expressada numa linguagem artificial formalmente definida; d) rigorosamente formal: termos definidos meticulosamente com semântica formal, teoremas e comprovação de propriedades como retidão e integralidade.
- Propósito: a intenção de uso da ontologia;
- Tema: a natureza do tema que a ontologia está caracterizando.

De acordo com os objetivos a que se propõem, as ontologias podem ser classificadas, segundo Gómez-Pérez, 2000, em:

- a) Ontologias de representação do conhecimento: compreendem as representações primitivas usadas para conhecimento formalizado em paradigmas de representação do conhecimento;
- b) Ontologias gerais/comuns: incluem vocabulários relacionados a coisas, eventos, tempo, espaço, causalidade, comportamento, função, etc.

- c) Meta-ontologias: também chamadas ontologias genéricas ou ontologias núcleo as quais são reutilizadas através dos domínios;
- d) Ontologias de domínio: são reutilizáveis em um determinado domínio. Elas fornecem vocabulários sobre os conceitos dentro de um domínio e seus relacionamentos, sobre as atividades que ocorrem em um domínio e sobre as teorias e princípios elementares reguladoras deste domínio;
- e) Ontologias de tarefa: fornecem um vocabulário sistematizado de termos para solução de problemas associados com tarefas que podem ou não ser do mesmo domínio. Estas ontologias fornecem um conjunto de termos por significados com as quais descrevem genericamente como resolver um tipo de problema;
- f) Ontologias de tarefas de domínio: são ontologias reutilizáveis em um dado domínio, mas não através do domínio;
- g) Ontologias de aplicação: contém o conhecimento necessário para modelagem de um domínio.

A utilização de ontologias como recurso nos SRI encerra algumas vantagens, conforme descreve Guimarães (2002): a) Elas fornecem vocabulário para representação do conhecimento apoiado na conceitualização, o quê reduz a ambigüidade; b) Permitem o compartilhamento do conhecimento; c) Fornecem descrição exata do conhecimento, uma vez que o vocabulário construído será amparado pela conceitualização comum dentro de um domínio; d) Uma mesma conceitualização pode ser expressa em várias línguas; e) Pode ser estendida de forma que se adeqüe a domínios específicos.

Por suas características e aplicabilidade, as ontologias podem ser utilizadas em áreas e para fins diversos, tais como:

- Processamento de linguagem natural: funcionando como um dicionário de termos apoiados nos conceitos, possibilita a eliminação das ambigüidades quando representa o conhecimento de um domínio;
- Gestão do conhecimento: possibilitam a estruturação básica para a elaboração de bases de dados e o processamento de dados não-estruturados;
- Web semântica: viabiliza buscas mais precisas, possibilitando a compreensão dos significados dos conteúdos;
- Comércio eletrônico: favorece a descrição dos produtos e a integração dos catálogos.

A possibilidade de aplicações variadas e as diversas classificações de ontologias demonstram que seu uso está diretamente relacionado à necessidade emergente de representação e compartilhamento do conhecimento e à diversidade dos SRI.

Nesse sentido, Chandrasekaran *et al.* (1999, p. 22) concluem que, apesar das diferenças, existe uma conformidade entre as ontologias em vários aspectos: a) Existem objetos no mundo; b) Os objetos têm propriedades ou atributos a serem valorizados; c) Os objetos podem existir em várias relações um com o outro; d) As propriedades e relações podem mudar no tempo; e) Há acontecimentos que ocorrem em instantes diferentes de tempo; f) Há processos em que objetos participam, e isso ocorre no tempo; g) O mundo e seus objetos podem estar em estados diferentes; h) Os acontecimentos podem causar outros acontecimentos como efeitos; i) Os objetos podem ter partes.

Além da representação do conhecimento, a possibilidade de reutilizar este conhecimento e expandir de forma a ser utilizado por outros domínios é o grande diferencial das ontologias. De igual importância, apresenta-se o fato dela contribuir para a redução da ambigüidade semântica que é característica da LN.

A utilização de ontologias como vocabulário de representação e recuperação do conhecimento vem ao encontro da necessidade de otimizar e favorecer a posição do usuário

frente às novas tecnologias e seus propósitos cada vez mais inovadores. A especificação do conceito, como propõe Gruber (1993), e a contextualização, indicada por Uschold (1996), são imprescindíveis para uma construção de ontologias que, sob o ponto de vista da recuperação na perspectiva do usuário, será considerada satisfatória.

CAPÍTULO 3 - CONTEXTO DE PESQUISA: O SISTEMA ORBIS_MC

3.1 APRESENTAÇÃO DO ORBIS_MC

O Observatório Regional Base de Indicadores de Sustentabilidade Metropolitano de Curitiba (ORBIS_MC) foi fundado em 31 de maio de 2004 e tem como objetivo “reunir indicadores de desenvolvimento urbano e de qualidade de vida dos 26 municípios da região”, conforme site institucional disponível em <<http://www.observatorio.org.br>>. Esses indicadores são coletados em fontes oficiais, depois avaliados por especialistas, para então seus resultados servirem de base para a formulação de políticas voltadas ao desenvolvimento uniforme da região.

Organismo integrante da Rede Mundial de Observatórios Urbanos, programa da ONU para Assentamentos Urbanos (UN-Habitat), conta com o apoio do Observatório Global Urbano (GUO), da Federação das Indústrias do Estado do Paraná (FIEP), Instituto Paraná Desenvolvimento (IPD), Governo do Estado por meio da Secretaria Especial de Assuntos Metropolitanos, Prefeitura de Curitiba, dentre outras instituições, que têm por objetivo ajudar a implantar a Agenda Habitat e a Agenda do Milênio (Agenda 21), gerando informações em âmbito local e nacional.

Os indicadores econômicos são utilizados para o acompanhamento das ações e permitem uma prospecção de futuro, haja vista que o ORBIS_MC se propõe apoiar as iniciativas governamental e privada de desenvolvimento sustentável para a Região Metropolitana de Curitiba (RMC).

Como base para o agrupamento e análise dos indicadores, o Observatório está apoiado na Agenda Habitat e nos Oito Objetivos do Milênio, estabelecidos na Declaração do Milênio

das Nações Unidas, em reunião da ONU com 147 Chefes de Estado e de Governo de 191 Países, em setembro de 2000. As metas do milênio levam em consideração questões como a erradicação da fome e da pobreza, educação, saúde e da sustentabilidade ambiental.

Conforme consta em sua divulgação institucional, “o Observatório selecionou parte dos 48 indicadores estabelecidos pela ONU, seguindo o critério de adequá-los à realidade da RMC, para serem aplicados na formulação de políticas e planejamento estratégico dos Municípios”. Baseados nisso, os especialistas e pesquisadores do Observatório efetuaram levantamento do perfil das localidades da região de abrangência, utilizando dados que demonstram a condição dos municípios e da vida da população. Com esses dados, foi elaborado o relatório “Indicadores do Milênio na Região Metropolitana de Curitiba” como forma de possibilitar a verificação dos resultados das ações da sociedade e dos governos para melhorar a qualidade de vida e a sustentabilidade da região.

3.2 APRESENTAÇÃO DO SISTEMA

O Sistema de Informações KMAI/ORBIS_MC tem o intuito de contribuir para a integração de instituições que compõem a RMC, além de estabelecer um ambiente cooperativo de troca de conhecimentos. A tecnologia KMAI® é utilizada para o monitoramento de informações em tempo real e para o cruzamento dessas informações. Como resultado, são produzidas informações estratégicas voltadas ao suporte dos processos de elaboração e tomada de decisão das políticas públicas.

Entre as principais vantagens do projeto, destacam-se: a) A indexação de documentos em tempo real nas fontes de informação monitoradas. Por meio desse mecanismo, o processo de captura, armazenamento e tratamento das informações torna-se mais ágil; b) Melhor aproveitamento de oportunidades e recursos em organismos de financiamento pelo

monitoramento constante das fontes de financiamento nacional e internacional; c) Visualização gráfica que facilita a interpretação dos fatos; d) Busca textual que facilita a recuperação de documentos relevantes ao conteúdo da consulta; f) Retro-alimentação do sistema, ou seja, a possibilidade de inserção de novos conhecimentos através de notas informativas que serão recuperadas pelo usuário. Nesse sentido, o aumento da utilização do sistema gera também um aumento de performance do mesmo. g) Por fim, a possibilidade de relações entre diversos aspectos; por exemplo: poluição do ar e camada de ozônio ou produção de recursos naturais e gestão do ambiente e de seus recursos.

O desenvolvimento do projeto permite a realização em paralelo da construção de uma cesta de indicadores, processo que está em sinergia com a produção de informações estratégicas geradas pelo sistema.

3.3 A FERRAMENTA KMAI®

A ferramenta KMAI® é composta das seguintes interfaces: Interface de Análise; Interface para inserção de Notas Informativas (as notas informativas são documentos de análise produzidos pela instituição); Interface de Relatório e Gráfica; e Interface de Administração.

Os seguintes módulos compõem a ferramenta:

- Análise: permite busca em linguagem aberta e resgata os documentos mais semelhantes com o texto de entrada. Utiliza técnicas de Inteligência Artificial e metodologias de recuperação da informação desenvolvidas no Brasil;
- Coleta: tem capacidade de reunião de informações digitais, advindas da Internet ou de arquivos locais, em uma base de conhecimento classificada e indexada de forma automática;

- Notas Informativas: insere conteúdos produzidos pelos analistas com possibilidade de recuperação posterior baseado em graus de similaridade;
- Suíte de Engenharia do Conhecimento: módulo destinado à definição dos conceitos e contextos (ontologias), que serão utilizados pelas rotinas de Inteligência Artificial para entenderem o universo dos assuntos;
- Relatórios Gráficos: visualização gráfica de todos os dados armazenados pelo sistema, classificados por tema, termos, fonte, período ou combinação destes atributos;
- Administração: disponibilização dos níveis de acesso ao sistema e gerenciamento da política de segurança.

3.4 RESULTADOS DO SISTEMA

O sistema KMAI/ORBIS_MC proporciona os seguintes resultados: 1) Monitoramento de assuntos específicos em tempo real; 2) Geração de gráficos; 3) Aumento do conhecimento sobre a RMC; 4) Fornecimento de subsídios para análises comparativas de diversos tipos de indicadores; 5) Aperfeiçoamento do modelo de gestão do conhecimento do ORBIS_MC.

3.5 A ELABORAÇÃO DO VOCABULÁRIO

O processo de construção de ontologias para utilização na tecnologia KMAI® se desenvolve em etapas diferenciadas, de acordo com a tipificação do documento-alvo de recuperação e análise.

Para a representação do conhecimento e posterior recuperação de documentos estruturados e semi-estruturados, o Sistema utiliza tecnologias desenvolvidas especialmente para este fim e que integram o Editor de Ontologias, chamados Extrator Semântico e Extrator

de Frequência. Trata-se de ferramenta que busca a identificação automática dos termos e está voltada à representação e recuperação em documentos internos próprios de cada área, estruturados em bases de dados próprias. Esse universo de documentos não será objeto de análise desta pesquisa. Nosso contexto se concentra nos documentos não-estruturados, descritos a seguir.

3.6 REPRESENTAÇÃO DO CONHECIMENTO EM DOCUMENTOS NÃO-ESTRUTURADOS

A construção de ontologias, para recuperação e análise de documentos não-estruturados e recuperados diretamente da Web, não conta com mecanismos automáticos, razão pela qual é vital a validação da qualidade das expressões construídas.

Nesse caso, onde os documentos não-estruturados são recuperados da mídia, identificada como fonte, o processo de elaboração das ontologias tem a participação efetiva dos especialistas e engenheiros do conhecimento. Infere-se que a identificação da fonte é essencial para a performance do Sistema, tendo relação direta com a construção das ontologias. Isto se dá porque as ontologias construídas devem, necessariamente, representar o conteúdo contido nas fontes. Outro fator importante está na representatividade da fonte que deve sempre conter qualidade nos seus conteúdos e ser de amplo conhecimento e utilização por parte da equipe de usuários.

A construção dos vocabulários está diretamente ligada à representação do conhecimento de um domínio. Assim, o conhecimento deste domínio precisa ser comum a todos os membros da equipe. Identifica-se que a participação do especialista é fundamental, uma vez que, além do conhecimento das estratégias e procedimentos usuais, é ele quem

fornece informações sobre o quê deve ser recuperado com eficiência pelo usuário e quais as particularidades de cada conteúdo.

Para que seja possível uma perfeita recuperação das informações, as linguagens utilizadas pelos usuários precisam ser “entendidas” pelo sistema de recuperação. O uso das ontologias busca favorecer este entendimento comum que tem seu ponto de partida na sincronização das equipes, na percepção e na elaboração dos conceitos (BUENO, 2005). Sendo uma forma de representação, a ontologia representa um entendimento comum, ou seja, tem o mesmo conceito para todos que utilizarem o sistema.

Nesta sincronização, a visualização do ambiente é fundamental, na qual todos os participantes das equipes devem ter a mesma compreensão do que está sendo representado e do que é objeto de recuperação. Nesta etapa, as dinâmicas como a percepção do ambiente, elementos envolvidos, rotinas, materiais utilizados, delimitação física, são alguns dos componentes utilizados para que todos os membros da equipe estabeleçam a mesma definição, tenham o mesmo conceito sobre o que vai ser representado e onde se dá o início da definição do sistema. Este processo de integração das equipes envolvidas no compartilhamento das informações vai contribuir para a definição dos termos e a construção dos relacionamentos necessários para a representação do conhecimento (BUENO, 2005).

Dessa maneira, o entendimento comum levará todos a perceber, por exemplo, que o termo “violência urbana” tem como relação conexa: segurança da população, segurança nas cidades, monitoramento das ruas. A partir desse exemplo, pode-se concluir que o conhecimento do ambiente vai permitir que todos identifiquem os aspectos e as ações envolvidas em cada área. Assim, a partir de uma definição do termo de maior abrangência, os demais vão sendo construídos conforme o seu relacionamento, formando as teias que no sistema estabelecem os relacionamentos semânticos.

3.7 CONSTRUÇÃO DO DICIONÁRIO DE TERMOS

O processo de identificação dos termos tem seu início quando da definição do contexto alvo. Esta definição consiste na percepção comum entre os participantes do processo de qual domínio será representado e quais os componentes ou ações que fazem parte deste. Isto significa que todos os envolvidos contribuem com a sua percepção do ambiente e do que será representado e qual a melhor forma de fazê-lo.

Na etapa seguinte, serão elaboradas as categorias, ou seja, a inclusão dos termos no universo do contexto e suas relações diretas ou indiretas. Essas relações, por sua vez, vão representar as interações dentro do domínio. Assim, após a visualização do ambiente e dos objetivos do sistema, é possível a identificação do vocabulário que será utilizado para a representação do domínio e que se constitui na fase seguinte do processo de construção.

Nesta identificação do vocabulário, são considerados os termos ou ainda as expressões utilizadas de forma corrente, isto é, utilizadas no dia-a-dia, além daquelas consideradas formais. Como formais, são classificados os termos técnicos ou aqueles encontrados em documentação oficial.

Esta distinção entre o vocabulário usual e o formal do domínio é necessária para a elaboração do dicionário, pois, seguindo os propósitos da ontologia, os termos e expressões elaborados devem traduzir, em linguagem natural e entendida por todos, o conhecimento que está contido em cada domínio.

Para a expansão do vocabulário, inicialmente, são considerados os sinônimos, uma vez que, a partir deles, é possível elaborar o entendimento do conteúdo e do contexto em que se pretende trabalhar. Com a lista de sinônimos, são considerados aqueles que têm representatividade na delimitação do contexto e que são de entendimento comum aos participantes. A elaboração inicial dos sinônimos representa também a condição de

caracterização do domínio, possibilitando aos engenheiros do conhecimento, que estão interagindo no processo com os especialistas, absorver e elaborar seu conhecimento prévio do que será representado.

Seguindo a metodologia adotada para a construção das ontologias de representação do conhecimento em documentos não-estruturados, a atividade de identificação do vocabulário conta com a participação dos especialistas do domínio, no sentido de apontar outros tipos de relacionamentos semânticos existentes entre os termos considerados expressivos para representar o domínio. Nesta identificação, todos contribuem num processo de *brainstorm*, em que os termos vão sendo elencados, de forma visível, para que todos possam ter a noção do contexto do domínio e os relacionamentos, além da pertinência de cada termo.

Na etapa de identificação do vocabulário, os membros das equipes podem também identificar se existem divergências de conceitualização, sincronizar e padronizar o entendimento dos termos elencados e do conhecimento que os termos pretendem representar, partilhamento do conhecimento para construção dos dicionários, além de ser o momento de sugestões e/ou padronização de vocabulário.

Segundo Bueno (2005, p. 117), a participação das Ontologias no sistema se dá em três momentos distintos:

- 1) Fase de Coleta: o sistema extrai as informações das bases que foram selecionadas pelo usuário. Os documentos são indexados individualmente com base nas Ontologias construídas pelos especialistas e engenheiros do conhecimento na fase de Engenharia do Conhecimento. O sistema então vai identificar os documentos com todas as expressões que forem localizadas no seu texto, armazenando-os de forma organizada na Base de Conhecimento.
- 2) Interfaces de análise disponibilizadas ao usuário final do sistema: o processo inicia quando o usuário digita o texto para pesquisa. Nesse momento, são identificadas as

Expressões Indicativas descritas pelo usuário que coincidem com aquelas previstas nas Ontologias. Essas expressões vão determinar o fluxo da cadeia de relações. O fluxo se estabelece de acordo com a classificação das Ontologias em: sinônimos, conexos, ‘tipo de’ e ‘parte de’. Verifica-se então que existe uma relação dinâmica entre a forma de como o usuário entra com uma Expressão Indicativa na interface de recuperação e a forma de como são definidas as relações no Editor de Ontologias. Essa dinamicidade está no fato de que quem determina a importância das expressões a serem consideradas é o próprio usuário, no momento da descrição do texto para pesquisa.

- 3) Editor de Ontologias: o Editor está integrado à arquitetura do sistema e fica disponível permitindo a atualização constante da base do conhecimento com novas expressões. Sempre que for feita uma alteração nas Ontologias, o sistema re-indexa todos os textos da base para garantir que as interfaces de análise possam trabalhar com todos os casos constantes da base de conhecimento.

CAPÍTULO 4 – METODOLOGIA

4.1 CARACTERIZAÇÃO DA PESQUISA

O estudo da História tem demonstrado que o conhecimento foi sendo construído ao longo dos tempos. Exaustivas pesquisas científicas levam a descobertas que se perpetuam e fomentam outras, criando, assim, o ciclo do desenvolvimento científico. Nessa construção, não apenas as teorias científicas são consideradas, mas também as práticas.

Pode-se dizer que o trabalho científico está apoiado em práticas que aliam técnica e pesquisa aplicada, métodos e metodologias de verificação, teoria e prática. Descartes, em seu *Discurso do Método* (s.d.), já dizia acreditar que poderia obter melhores raciocínios e conclusões daqueles que estão na prática desenvolvendo seus trabalhos, do que acompanhando apenas quem produz teorias e proposições.

Nesse esteio, a presente pesquisa apoiou-se no estudo de publicações já reconhecidas nas áreas de Ciência da Informação e Lingüística, voltando-se para uma aplicação e análise prática. Pode-se delimitá-la como do tipo qualitativo e bibliográfico, com aplicação em estudo de caso.

Portanto, este estudo é caracterizado como tendo uma abordagem qualitativa, pois, segundo Goldenberg (2004, p. 14), na pesquisa qualitativa, “a preocupação do pesquisador não é com a representatividade numérica do grupo pesquisado, mas com o aprofundamento da compreensão de um grupo social, de uma organização, de uma instituição, de uma trajetória”. Para Flick (2004, p. 27),

[...] o processo de pesquisa qualitativa pode ser representado como uma trajetória que parte da teoria em direção ao texto, e outra do texto de volta para a teoria. A interseção dessas duas trajetórias é a coleta de dados verbais ou visuais e a interpretação destes em um plano de pesquisa específico.

O autor considera ainda que (2004, p. 28), “[...] A pesquisa qualitativa é orientada para a análise de casos concretos em sua particularidade temporal e local, partindo das expressões e atividades das pessoas em seus contextos locais”.

Dentre os aspectos que envolvem os métodos qualitativos, propostos por Tobar (2004, p. 92), destacam-se: 1) Ocupar-se mais dos significados do que da frequência dos fatos; 2) Operar com símbolos lingüísticos, por exemplo: com as metáforas e conceitos, com as narrativas e descrições, mais do que com números e estatísticas.

Quanto ao aspecto bibliográfico, na seleção da literatura, buscou-se identificar documentos que versassem sobre a utilização de linguagem natural e suas novas formas de representação, objetivando especificamente estudar as ontologias. Dessa análise, foram sintetizadas recomendações sobre o estabelecimento das relações semânticas entre os termos com vistas a definir critérios de validação de ontologias sob o ponto de vista lingüístico.

Os critérios foram aplicados em uma amostra da ontologia utilizada na tecnologia KMAI® do ORBIS_MC. Tal Sistema foi desenvolvido com base neste recurso de linguagem como forma de recuperação da informação em documentos não-estruturados, disponíveis em mídias escritas, tais como: jornais, revistas e sites eletrônicos com código aberto.

A aplicação dos critérios semânticos na amostra da ontologia utilizou técnicas de análise de conteúdo, propostas por Bardin (2004, p. 37), que consiste em:

[...] Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos sistemáticos e objectivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens.

Para a autora, a análise de conteúdo vai estudar “a prática da língua por emissores identificáveis”, além de considerar “as significações (conteúdo)”. A análise de conteúdo procura conhecer aquilo que está por trás das palavras sobre as quais se debruça. É uma busca de outras realidades através das mensagens (BARDIN, 2004, p. 38).

A utilização da técnica de análise do conteúdo ocorreu seguindo as três etapas propostas pela autora: 1) a pré-análise; 2) a exploração do material; 3) tratamento dos resultados, a inferência e a interpretação (BARDIN, 2004, p. 89).

A seguir, descreve-se cada uma delas, especificando-as no objeto desta pesquisa.

4.2 COLETA DE DADOS (PRÉ-ANÁLISE)

A coleta dos dados foi realizada no dicionário que está instalado no Sistema KMAI ®. Esse dicionário foi construído por um grupo de especialistas com o objetivo de monitorar a situação da Região Metropolitana de Curitiba em três aspectos, que determinaram o escopo do Sistema: Saúde Ambiental, Saúde Humana e Violência Urbana. O dicionário é composto por expressões indicativas, e suas relações de sinônimos (equivalência), ‘tipo de’ e/ou ‘parte de’ (hierárquicas), e conexos (associativas), foram organizados em três domínios e 15 subdomínios, conforme segue:

DOMÍNIOS	SUBDOMÍNIOS
Saúde Ambiental:	Energia; Gestão de Recursos Naturais; Saneamento Ambiental; Poluição Ambiental; Ocupação do Solo; Órgãos vinculados à Saúde Ambiental.
Saúde Humana na RMC:	Promoção da Saúde; Vigilância em Saúde; Políticas de Saúde; Gestão de Saúde.
Violência Urbana na RMC	Modalidades de Violência Urbana; Violência baseada por localidade; Causas da Violência Urbana; Consequências da Violência Urbana; Segurança Pública

Quadro 1: Domínios e subdomínios

Entende-se por ‘domínio’ como a área em que se pretende trabalhar ou a definição do universo que será representado. O ‘subdomínio’ é a divisão do domínio nas temáticas suficientes para delimitar o assunto ou, ainda, representar o conteúdo pretendido. Os termos ou expressões indicativas são as palavras ou o conjunto de palavras que expressam um significado, representando o domínio de aplicação do sistema.

Com base nos critérios de exaustividade e pertinência, propostos por Bardin (2004, p.90-92), a amostra de ontologias contemplou duas expressões de cada subdomínio e suas relações semânticas. As expressões foram escolhidas aleatoriamente em cada subdomínio, sendo que, para melhor representar o conteúdo e proceder a análise, estabeleceu-se como critério a escolha de expressões com mais de quatro relações semânticas dentro dos três tipos propostos: equivalência, hierárquicas e associativas. Ressalta-se, porém, que, no universo analisado, cinco expressões geraram apenas duas relações semânticas, sendo consideradas, por não haver outras dentro do seu domínio, com mais de quatro relações. Também foi encontrado um subdomínio contendo termos com apenas três ou menos relações semânticas.

Como resultado, obteve-se uma base de conhecimento para análise composta de: três domínios; 15 subdomínios; 28 termos ou expressões indicativas; e 163 relações semânticas construídas. Registra-se que o subdomínio “Órgãos Vinculados a Saúde Ambiental” não apresenta nenhuma expressão.

O universo a ser analisado se apresenta conforme descrito a seguir:

DOMÍNIO: SAÚDE AMBIENTAL NA REGIÃO METROPOLITANA DE CURITIBA		
SUBDOMÍNIO: ENERGIA		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Biodiesel	Combustível natural	sinônimo
	Combustível biológico	sinônimo
	Biocombustível	sinônimo
	Pesquisa sobre biocombustível	conexo
	Aumento da produção de biodiesel	conexo
	Comercialização de biodiesel	conexo
	Óleo biocombustível ecológico	tipo de
Combustível biológico	Biodiesel	sinônimo
	Combustível natural	sinônimo
	Biocombustível	sinônimo
	Pesquisa sobre biocombustível	conexo
	Aumento da produção de biodiesel	conexo
	Comercialização de biodiesel	conexo
	Óleo biocombustível ecológico	tipo de

Quadro 2: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Energia

DOMÍNIO: SAÚDE AMBIENTAL NA REGIÃO METROPOLITANA DE CURITIBA		
SUBDOMÍNIO: GESTÃO DE RECURSOS NATURAIS		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Áreas Verdes	Zonas verdes	sinônimo
	Espaços verdes	sinônimo
	Recuperação de espaços verdes	conexo
	Espaços verdes de proteção	conexo
	Gestão de espaços verdes	conexo
	Regulamentação de espaços verdes	conexo
	Rede de espaços verdes	parte de
	Espaços verdes integrados	tipo de
Manejo Ambiental	Manejo de recurso natural	sinônimo
	Gestão do Ambiente e de seus recursos	sinônimo
	Gerenciamento de recursos naturais	sinônimo
	Gestão de recurso natural	sinônimo
	Produção de recursos naturais	conexo
	Políticas de gestão de recursos naturais	conexo

Quadro 3: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Gestão de Recursos Naturais

DOMÍNIO: SAÚDE AMBIENTAL NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: SANEAMENTO AMBIENTAL		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Aterro sanitário	Depósitos de resíduos sólidos urbanos	sinônimo
	Depósitos de lixo	sinônimo
	Fechamento de aterros sanitários	conexo
	Saturamento do aterro sanitário	conexo
	Aterro sanitário controlado	conexo
Coleta e tratamento de Resíduos	Tratamento de resíduos	sinônimo
	Empresas de tratamento de resíduos	conexo
	Central de tratamento de resíduos	conexo
	Sistemas de tratamento de resíduos	conexo
	Tratamento de resíduos químicos	tipo de
	Tratamento de resíduos sanitários	tipo de
	Tratamento de resíduos sólidos	tipo de
	Tratamento de águas residuárias	tipo de
	Tratamento de resíduos industriais	tipo de
	Tratamento de resíduos orgânicos	tipo de

Quadro 4: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Saneamento Ambiental

DOMÍNIO: SAÚDE AMBIENTAL NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: POLUIÇÃO AMBIENTAL		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Área contaminada	Poluição do solo	sinônimo
	Degradação do solo	sinônimo
	Contaminação do solo	sinônimo
	Grau de contaminação do solo	conexo
	Situação de contaminação do solo	conexo
	Nível de contaminação do solo	conexo
	Processo de contaminação do solo	conexo
	Aumento da degradação do solo	conexo
	Redução da degradação do solo	conexo
	Controle da degradação do solo	conexo
Destruição dos Recursos Hídricos	Poluição hídrica	sinônimo
	Degradação de recursos hídricos	sinônimo
	Degradação dos rios	sinônimo
	Deterioração dos rios	sinônimo
	Destruição dos rios	sinônimo
	Deterioração de recursos hídricos	sinônimo
	Contaminação de recursos hídricos	sinônimo
	Poluição de recursos hídricos	sinônimo
	Eutrofização de lagos	conexo
	Salinização de rios e açudes	conexo
	Doenças ligadas à poluição hídrica	conexo

Quadro 5: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Poluição Ambiental

DOMÍNIO: SAÚDE AMBIENTAL NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: OCUPAÇÃO DO SOLO		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Explosão demográfica urbana	Crescimento da população	sinônimo
	Expansão populacional	sinônimo
	Aumento populacional nas cidades	sinônimo
	Crescimento populacional	sinônimo
	Crescimento populacional desordenado	conexo
Loteamento irregular	Loteamento clandestino	sinônimo
	Regularização de loteamentos irregulares	conexo
	Regularização de loteamentos clandestinos	conexo
	Proliferação de loteamentos clandestinos	conexo

Quadro 6: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Ocupação do Solo

DOMÍNIO: SAÚDE AMBIENTAL NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: ÓRGÃOS VINCULADOS À SAÚDE AMBIENTAL		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
-	Nenhuma relação encontrada	-

Quadro 7: Universo de termos e relações analisados. Domínio: Saúde Ambiental. Subdomínio: Órgãos vinculados à Saúde Ambiental

DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: PROMOÇÃO DA SAÚDE		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Ações de educação em saúde	Práticas de educação em saúde	sinônimo
	Atividades de educação em saúde	sinônimo
	Educação em saúde	sinônimo
Prevenção da saúde	Questão de saúde	parte de
	Questão de saúde pública	parte de
	Educação e prevenção em saúde	tipo de
	Promoção e prevenção em saúde	tipo de
	Serviço de prevenção de saúde	tipo de
	Programa de saúde pública	tipo de
	Prevenção de doença	tipo de

Quadro 8: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Promoção da Saúde

DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: VIGILÂNCIA EM SAÚDE		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Aumento de doenças	Casos de aumento de enfermidades	sinônimo
	Surto de doenças	sinônimo
	Surto de epidemias	sinônimo
	Aumento de epidemias	sinônimo
	Aumento de enfermidades	sinônimo
	Aumento de enfermidade grave	conexo
	Aumento de enfermidades transmissíveis	conexo
Falta de condição de higiene	Falta de saneamento	conexo
	Condições de higiene	conexo
	Falta de saneamento básico	conexo
	Origem de doenças	tipo de
	Motivo de doenças	tipo de
	Causa de enfermidades	tipo de
	Agravo em saúde	tipo de
	Causa de doenças	tipo de

Quadro 9: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Vigilância em Saúde

DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: POLÍTICAS DE SAÚDE		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Ações de saúde	Campanha de saúde	sinônimo
	Programas de saúde	sinônimo
	Implantação de programa de saúde	conexo
	Falta de programa de saúde	conexo
	Execução de programa de saúde	conexo
	Planejamento de programa de saúde	conexo
	Ações de saúde coletiva	tipo de
	Ações de saúde pública	tipo de
Campanha de saúde	Ações de saúde	sinônimo
	Programas de saúde	sinônimo
	Falta de programa de saúde	conexo
	Execução de programa de saúde	conexo
	Planejamento de programa de saúde	conexo
	Implantação de programa de saúde	conexo
	Campanha de saúde pública	conexo
	Ações de saúde coletiva	tipo de
	Ações de saúde pública	tipo de

Quadro 10: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Políticas de Saúde

DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: GESTÃO DE SAÚDE		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Custo de saúde	Aplicação de recurso na saúde	sinônimo
	Orçamento da saúde	sinônimo
	Investimento em saúde	sinônimo
	Aumento do investimento em saúde	conexo
	Redução do investimento em saúde	conexo
	Recuperação de investimento em saúde	conexo
Investimento em saúde	Orçamento da saúde	sinônimo
	Custo da saúde	sinônimo
	Aplicação de recurso na saúde	sinônimo
	Aumento do investimento em saúde	conexo
	Redução do investimento em saúde	conexo
	Recuperação de investimento em saúde	conexo

Quadro 11: Universo de termos e relações analisados. Domínio: Saúde Humana. Subdomínio: Gestão de Saúde

DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: MODALIDADES DE VIOLÊNCIA URBANA		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Violência nos presídios	Violência prisional	sinônimo
	Rebeliões prisionais	conexo
	Aumento da violência nos presídios	conexo
	Tortura de detentos	tipo de
	Maus-tratos aos detentos	tipo de
Violência no trânsito	Problema da violência no trânsito	sinônimo
	Combate à violência no trânsito	conexo
	Roubo no trânsito	tipo de
	Seqüestros no trânsito	tipo de
	Homicídios no trânsito	tipo de
	Assassinatos no trânsito	tipo de
	Assalto no trânsito	tipo de

Quadro 12: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Modalidades de Violência Urbana

DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: VIOLÊNCIA BASEADA POR LOCALIDADE		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Violência doméstica	Violência no lar	sinônimo
	Violência familiar	sinônimo
	Agressão doméstica	sinônimo
	Violência domiciliar	sinônimo
	Agressão familiar	sinônimo
Violência na periferia	Violência no bairro	sinônimo
	Violência na comunidade	sinônimo

Quadro 13: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Violência baseada por Localidade

DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: CAUSAS DA VIOLÊNCIA URBANA		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Gangues juvenis	Gangues nas escolas	sinônimo
	Vandalismo nas escolas	conexo
	Adolescentes infratores	parte de
	Menores infratores	parte de
Tráfico de armas	Tiroteio entre traficantes	conexo
	Controle do tráfico de armas	conexo
	Combate ao tráfico de armas	conexo
	Tiroteio entre quadrilhas	conexo

Quadro 14: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Causas da Violência Urbana

DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: CONSEQUÊNCIAS DA VIOLÊNCIA URBANA		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Depredação de áreas públicas	Depredação de equipamentos públicos	sinônimo
	Destruição de áreas públicas	sinônimo
Mortalidade por homicídio	Aumento de homicídios	conexo
	Aumento do número de homicídios	conexo
	Redução do número de homicídios	conexo

Quadro 15: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Consequências da Violência Urbana

DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA		
SUB DOMÍNIO: SEGURANÇA PÚBLICA		
TERMO	RELAÇÃO	TIPO DE RELAÇÃO
Custos da violência	Custos da criminalidade	sinônimo
	Custos econômicos da violência	sinônimo
Policiamento comunitário	Parceria da comunidade com a polícia	sinônimo
	Participação da comunidade na segurança	sinônimo

Quadro 16: Universo de termos e relações analisados. Domínio: Violência Urbana. Subdomínio: Segurança Pública

4.3 EXPLORAÇÃO DO MATERIAL

Como ponto de partida, a pesquisa identificou, pela análise da literatura de Ciência da Informação e Linguística, as características conceituais que determinam direta ou indiretamente as relações de equivalência, hierárquicas e associativas. Essas indicações teóricas são amplamente utilizadas na construção de tesauros e constatou-se que elas são perfeitamente aplicáveis na construção de ontologias devido às semelhanças nos processos de construção.

Dessa análise, foram estabelecidas as relações de equivalência, hierárquicas e associativas entre os autores e os critérios que serão considerados:

RELAÇÕES DE EQUIVALÊNCIA - SINONÍMIA	
ISO 704:2000	Relação entre designações diferentes que designa o mesmo conceito, tendo a mesma intensão. São intercambiáveis.
ISO/R 1087-1969	Termos que têm o mesmo significado – exatamente ou próximo – mas com diferentes formas externas.
Gramática da Língua Portuguesa	São estabelecidas entre duas palavras ou mais que apresentam significados iguais ou semelhantes.
Lyons (1987)	Se tiverem o mesmo significado descritivo, expressivo e social (na faixa de contextos em questão).
Cintra (2002)	Identidade de significação entre elementos lexicais.
Dubois et.al. (1986)	Quando os termos têm a possibilidade de se substituírem um ao outro ou quando dois termos são intercambiáveis em todos os contextos. Tendência das unidades do léxico de terem o mesmo significado e de serem substituíveis umas pelas outras.
Dahlberg (1978)	Relações lógicas de identidade, onde as características dos conceitos são as mesmas.
Faulstich (2005)	Quando considerados dois termos, um pode substituir o outro em determinado enunciado. Quando dois termos são intercambiáveis em todos os contextos.
IBICT (1984)	Quando dois ou mais termos são considerados, para fins de indexação, como designando o mesmo conceito.

Quadro 17: Síntese de Autores e Critérios. Relações de Equivalência – Sinonímia

RELAÇÕES HIERÁRQUICAS – TIPO DE / PARTE DE	
Gomes (1990)	Ordenação dos termos determina uma verticalização: termos superordenados (com todas as características) se ligam a termos subordinados (com pelo menos uma característica a mais).
Cintra (2002)	Relações gênericas: envolvem e compartilham amplamente todas as características dos termos: têm características comuns. Relações específicas: apresentam pelo menos uma característica diferente das características genéricas. Relações partitivas: envolvem características de um objeto como um todo e seus componentes como parte.
ISO 1087	Definem-se nas noções subordinadas em um ou vários níveis. Caracterizada como gênero e espécie, são estabelecidas quando dois conceitos apresentam características semelhantes, porém, um tem uma característica a mais que o outro.
Sager (1990)	As relações partitivas servem para indicar a conexão entre conceitos consistindo num todo e suas partes.
ISO 704/2000	Quando o conceito superordenado representa um contexto maior, total, enquanto os conceitos relacionados ou inferiores representam as partes deste total.
Cabré (1993)	As relações partitivas são relações de coordenação, onde o todo e suas partes se relacionam. Baseiam-se na contigüidade dos objetos no espaço, ou seja, simultaneamente.

Quadro 18: Síntese de Autores e Critérios. Relações Hierárquicas – Tipo de / Parte de

RELAÇÕES ASSOCIATIVAS – CONEXOS	
IBICT (1984)	Termos associados mentalmente onde há uma ligação explícita de subordinação, porém não são equivalentes e nem formam hierarquia.
ISO 704/2000	Ligação temática entre conceitos, mas não hierárquicas.
ISO 2788/1974	Normalmente utilizada para as outras relações que não são equivalentes nem hierárquicas.
Marroni (2006)	Ação, Coisa, Conceito, Processo ou Propriedade / Atributo, Princípio ou Propriedade
	Ação, Coisa, Matéria-prima ou Processo / Aplicação, Produto, Produto ou Resultado
	Ação, Conceito ou Processo / Agente (Instrumento) ou Paciente (Alvo)
	Causa / Efeito
	Coisa ou Conceito / Origem
	Coisa ou Processo / Estado ou Local
	Coisa ou Todo / Parte
	Conceito / Mecanismos de medida
	Disciplina, Campo de estudo ou Ocupação / Objeto ou fenômenos estudados ou aqueles que a praticam (Relação material)
	Expressões ou Frases sincategoremáticas / Substantivos nelas incluídos
	Oposição

Quadro 19: Síntese de Autores e Critérios. Relações Associativas – Conexos

Para a verificação das relações associativas, ressalta-se que foram utilizados principalmente os parâmetros sintetizados e propostos por Marroni (2006). A opção pela autora está diretamente ligada ao propósito desta pesquisa: avaliação e identificação de critérios e argumentos válidos para a construção de uma metodologia de apoio para construção de futuras relações. A sistematização dos parâmetros para o estabelecimento de relações associativas favorece o entendimento de como estas relações podem ocorrer no sistema.

Por essa razão, para a identificação das adequações ou inconsistências na construção do instrumento de representação do conhecimento, é importante relacionar a definição dos conceitos que compõem os parâmetros das relações associativas apresentada por Marroni (2006, p. 90), conforme o Quadro 20:

RELAÇÕES ASSOCIATIVAS	CONCEITOS
AÇÃO	Atuação, ato, efeito, maneira como um agente atua sobre outro.
AGENTE	Pessoa que pratica a ação.
ALVO	Aquilo que se procura alcançar, objetivo.
APLICAÇÃO	Emprego, utilização, uso.
ATRIBUTO	Aquilo que é próprio do ser.
CAUSA	Razão de ser, o que faz com que algo exista.
COISA	Aquilo que existe ou pode existir, objeto inanimado.
CONCEITO	Representação de um objeto pelo pensamento, por meio de suas características gerais.
DISCIPLINA	Qualquer ramo do conhecimento (artístico, científico, histórico, etc).
EFEITO	Consequência, resultado.
ESTADO	Situação ou disposição em que se acham as pessoas ou as coisas em um momento dado.
FENÔMENO	Fato, aspecto ou ocorrência passível de observação.
INSTRUMENTO	Recurso empregado para se alcançar um objetivo, conseguir um resultado; meio.
LOCAL	Area de limites definidos.
MATÉRIA-PRIMA:	Substância bruta principal com que é fabricada alguma coisa.
MECANISMOS DE MEDIDA	Mecanismo (conjunto de elementos que concorrem para a atividade de uma estrutura) + Medida (quantidade fixada por um padrão para determinar as dimensões ou o valor de uma grandeza da mesma espécie).
OBJETO	Tudo que é perceptível por qualquer dos sentidos.
OCUPAÇÃO	Atividade, serviço ou trabalho manual ou intelectual realizado por um período de tempo mais ou menos longo.
OPOSIÇÃO	Antagonismo, contraste
ORIGEM	Princípio, começo, procedência.
PACIENTE	O que sofre ou é objeto de uma ação.
PARTE	Porção de um todo dividido.
PRINCÍPIO	Preceito, regra, lei.
PROCESSO	Seqüência contínua de fatos ou operações que apresentam certa unidade ou que se reproduzem com certa regularidade.
PRODUTO	Aquilo que resulta de qualquer processo ou atividade.
PROPRIEDADE	Qualidade inerente aos seres.
RESULTADO	Consequência, efeito, termo, fim.
SINCATEGOREMÁTICO	Vocabulário que não tem significado por si mesmo, somente quando acompanhado de outros.
TODO	Completo, inteiro, total.

Quadro 20: Definição dos conceitos que compõem os parâmetros das relações associativas apresentada por Marroni (2006).

Como passo seguinte, os termos do dicionário e suas relações semânticas foram analisados individualmente. Para tal, elaborou-se uma comparação entre os critérios identificados na literatura e as relações semânticas estabelecidas para cada expressão indicativa. Nessa comparação, foram verificados as possibilidades de relações de equivalência, os tipos de relações hierárquicas e suas características e os parâmetros de relacionamento de cada relação associativa construída.

Na análise individual das relações, foi considerada também a pertinência da classificação de domínios e subdomínios a que estão submetidas. Com base nessa análise, foram identificadas as adequações e as inconsistências das relações.

A seguir, descreve-se a etapa do tratamento dos resultados que, devido a sua natureza e importância, se constitui num capítulo à parte.

CAPÍTULO 5 - TRATAMENTO DOS RESULTADOS

5.1 ANÁLISE QUANTITATIVA DAS RELAÇÕES

Após a seleção dos termos do dicionário, estes foram agrupados, segundo os três tipos de relação semântica proposta para análise na pesquisa: Sinonímicas; Hierárquicas; e Conexos (associativas).

O arranjo foi feito dispondo em uma tabela os termos ou expressões indicativas extraídos do dicionário do sistema em uma coluna, e as relações semânticas correspondentes dispostas em outra coluna, conforme Anexo 1. A disposição seguiu a ordem sequencial apresentada no dicionário, não obedecendo a critérios de alfabetação nem qualquer outro critério de relevância entre os termos. Tampouco os termos foram ordenados numericamente.

A tabela contém ainda duas colunas à direita para registro de dados sobre: a) Indicação de adequação ou inconsistência da relação semântica atribuída na ontologia; b) Comentários sugerindo correção da relação semântica atribuída ou oferecendo informação esclarecedora.

Este agrupamento inicial buscou permitir a visualização por tipo de relação do universo a ser analisado. A Tabela 1 mostra uma síntese quantitativa dos três tipos de relações semânticas estudadas:

Tabela 1: Síntese quantitativa das relações semânticas

TIPO DE RELAÇÕES	QUANTIDADE
1) SINÔNIMOS	65
2) HIERÁRQUICAS	
2.1) Tipo de	30
2.2) Parte de	05
3) ASSOCIATIVAS (CONEXOS)	63
TOTAL	163

Analisando a Tabela 1, é possível observar que as relações de sinônimos e conexos representam a maioria das relações construídas. Essas relações são consideradas as mais

importantes e de maior valor pelo Sistema KMAI®, conforme indicado pelos pesos atribuídos para o cálculo da similaridade apresentada na Tabela 2:

Tabela 2: Pesos para o cálculo de similaridade

RELAÇÃO DA SUÍTE DE ENGENHARIA DO CONHECIMENTO	PESOS PARA O CÁLCULO DA SIMILARIDADE
Sinônimos	0,99
Tipo de (hipernímia e hiponímia)	0,4
Parte de (meronímia e holonímia)	0,3
Termos Conexos	0,7

Fonte: BUENO (2004, p. 120)

Conforme identificado *in loco* e indicado na documentação do sistema, a verificação de similaridade entre os termos individuais e as expressões armazenadas na base de conhecimento é realizada por uma métrica. A métrica de similaridade no Sistema KMAI ® se dá por cálculo realizado para identificar a semelhança entre o conteúdo pesquisado e os documentos que estão na base do sistema. Para efeito de cálculo, são considerados os pesos das relações estabelecidas, conforme Tabela 2.

Como ponto de partida, buscaram-se os sinônimos para a determinação das relações. Constatou-se, nesta opção, que ela está apoiada na tentativa de identificar, em Linguagem Natural, palavras com o mesmo sentido num mesmo contexto, objetivando estabelecer um amplo universo de representações do conhecimento alvo e reduzir a pluralidade de designações encontradas, favorecendo, assim, a recuperação da informação.

As relações associativas, ou conexos como são identificadas no Sistema, são definidas após a construção dos sinônimos. Essas relações, por suas características, tendem a abranger um grande volume das informações a serem identificadas, recuperadas e indexadas para formação da base de conhecimento. Por se tratarem de relações complexas e encerrarem importância considerável na concepção dos mecanismos que permitem a recuperação dos conteúdos desejados, as relações associativas foram alvo de estudo mais detalhado nesta pesquisa.

Analisando a construção do dicionário incluso no Sistema, verificou-se que a elaboração dos termos conexos está diretamente relacionada ao conhecimento do especialista no domínio, que irá identificar as conexões entre os assuntos. Outro recurso utilizado foi a localização de expressões que aparecem com maior frequência em documentos que tratam do mesmo assunto.

As relações hierárquicas somente são utilizadas quando é imprescindível, ao objetivo do Sistema, o detalhamento e representação de todas as partes e tipos de um conceito. Baseado nesta premissa, verificou-se que a valoração dessa relação representa peso menor do que a relação de sinônimos e de conexos.

A exploração do *corpus* extraído do dicionário resultou na identificação de algumas relações semânticas inadequadas, segundo os critérios adotados com base na literatura. A Tabela 3 mostra, de forma quantitativa, o total de relações adequadas e inconsistentes verificadas no *corpus*.

Tabela 3: Síntese qualitativa das relações semânticas

RELAÇÃO	ADEQUAÇÃO	INCONSISTÊNCIA
SINÔNIMO	50	15
HIERÁRQUICA	24	11
ASSOCIATIVA	60	03
TOTAL	134	29

Durante a análise que resultou na compilação dos dados expressos na Tabela acima, foi possível verificar que as construções das relações nem sempre seguem um método que garanta a sua consistência. Este fato decorre, provavelmente, de um desconhecimento por parte do especialista dos rigores metodológicos necessários em atividades dessa natureza.

A seção a seguir apresenta, em maior detalhe, considerações de cunho qualitativo a respeito desses dados.

5.2 ANÁLISE QUALITATIVA DAS RELAÇÕES

A análise dos termos e suas relações estão apresentadas em dois momentos distintos. Inicialmente, são tratadas as relações que apresentam adequação aos critérios lingüísticos identificados na literatura. Em seguida, são descritas as relações consideradas inadequadas segundo esses mesmos critérios.

5.2.1 Relações adequadas

As relações de sinônimo, que apresentaram adequação na sua construção, correspondem às características que indicam relações em que os termos têm o mesmo conceito, sendo intercambiáveis em todos os contextos, conforme propõe Faulstich (2005) e Lyons (1987). Verificou-se que essas características foram encontradas em todas as ocorrências, considerando que os termos relacionados expressam um mesmo conceito dentro do domínio que pretendem representar.

Os exemplos a seguir expressam esta idéia:

Domínio: Saúde Ambiental

Subdomínio: Energia

biodiesel ↔ combustível natural

Domínio: Saúde Ambiental

Subdomínio: Gestão de Recursos Naturais

manejo ambiental ↔ manejo de recurso natural

Domínio: Saúde Ambiental

Subdomínio: Gestão de Recursos Naturais

áreas verdes ↔ espaços verdes

Domínio: Violência Urbana na Região Metropolitana de Curitiba

Subdomínio: Modalidades de Violência Urbana

violência doméstica ↔ agressão familiar

Domínio: Saúde Ambiental

Subdomínio: Ocupação do Solo

explosão demográfica urbana ↔ aumento populacional nas cidades

Foi possível verificar que, respeitando o domínio e o subdomínio em que estão classificados, os termos podem ser substituídos mutuamente no contexto, mantendo o entendimento e a recuperação das informações pretendidas.

As relações hierárquicas se apresentam em menor número no sistema. Isto ocorre pela metodologia adotada para construção da ontologia. Ela parte do estabelecimento de sinônimos e conexos na expectativa de abranger conteúdo bastante significativo no que tange à quantidade de representações do conhecimento passíveis de serem recuperadas. As relações hierárquicas são construídas a partir de situações específicas encontradas nas fontes para construção das ontologias.

Nessas relações hierárquicas, observou-se o estabelecimento de cadeias verticalizadas, com termos superordenados e subordinados bem definidos e com características comuns que justificam a relação, conforme exemplos a seguir:

Domínio: Saúde Ambiental

Sub Domínio: Saneamento Ambiental

coleta e tratamento de resíduos \supset ¹⁶tratamento de resíduos químicos

Domínio: Saúde Humana na Região Metropolitana de Curitiba

Sub Domínio: Políticas de Saúde

ações de saúde \supset ações de saúde pública

Domínio: Violência Urbana na Região Metropolitana de Curitiba

Sub Domínio: Modalidades de Violência Urbana

violência nos presídios \supset tortura de detentos

Domínio: Violência Urbana na Região Metropolitana de Curitiba

Sub Domínio: Modalidades de Violência Urbana

violência no trânsito \supset assalto no trânsito

Foi possível constatar que as relações se enquadram quase na sua totalidade no tipo genérico/específico, o que garante um detalhamento maior do universo de representação do conhecimento.

Na análise das relações hierárquicas, registrou-se uma situação em que duas relações estabelecidas poderiam ser classificadas como associativas:

¹⁶ \supset = contém

Domínio: Violência Urbana na Região Metropolitana de Curitiba

Sub Domínio: Causas da Violência Urbana

gângues juvenis \supset adolescentes infratores

gângues juvenis \supset menores infratores

Nesse caso, verificou-se a possibilidade da relação ‘parte de’ como do tipo associativa, conforme descrito por Marroni, 2006. No entanto, considerou-se adequada, pois a relação todo/parte é também um tipo de relação hierárquica.

As relações associativas foram analisadas em duas etapas. Inicialmente, buscou-se apurar se as relações construídas não se enquadravam nas categorias de sinonímica ou hierárquica. Isto porque, por se tratarem de relações de certa subjetividade, a tendência é indicar qualquer relação que, aparentemente, não tenha uma caracterização bem definida (sinônimo e/ou hierárquica) como conexo. Esta também é uma definição indicada por muitos autores, como Gomes (1984).

No resultado, foram identificadas três relações inadequadas de acordo com os critérios propostos. Essas relações serão descritas e comentadas na próxima seção.

Após a primeira análise, todas as relações associativas foram classificadas de acordo com os dez parâmetros sugeridos por Marroni (2006) e adotados como referência para esta pesquisa. Vale ressaltar que, nesta etapa, foi considerado o princípio da reciprocidade entre os termos. A reciprocidade favorece a análise, permitindo a troca de posição dos termos.

Dessa maneira, foi possível constatar que as relações associativas estudadas apresentaram relações com parâmetros diferentes, quando um mesmo termo estava associado a termos diversos, conforme ilustra o exemplo a seguir:

Domínio: Violência Urbana na Região Metropolitana de Curitiba

Sub Domínio: Causas da Violência Urbana

tráfico de armas \equiv tiroteio entre traficantes (**causa / efeito**)

tráfico de armas \equiv controle do tráfico de armas (**ação / alvo**)

tráfico de armas \equiv combate ao tráfico de armas (**ação / alvo**)

tráfico de armas \equiv tiroteio entre quadrilhas (**causa / efeito**)

Esta construção demonstra que, dentro de um domínio, busca-se representar o conhecimento de forma bastante ampla, registrando vários tipos de associações. O recurso da relação associativa cumpre o papel de ampliar a pesquisa na recuperação da informação.

5.2.2 Relações inconsistentes

No processo de análise, identificou-se a existência de 29 relações não adequadas às categorias em que estavam classificadas. Essas relações apresentaram características divergentes ou insuficientes que justificassem a classificação original. As inconsistências verificadas em cada tipo de relação semântica, propondo a solução correspondente, serão tratadas na sequência.

5.2.2.1 Relações classificadas originalmente como Sinônimos

As relações examinadas não se caracterizam como sinonímicas, pois não consideram o critério de identidade definido por Dahlberg (1978) e o requisito de intercâmbio de contexto constatado por Faulstich (2005) e Lyons (1987). Para Dahlberg (1978), nas relações lógicas de identidade que definem os sinônimos, as propriedades dos conceitos de cada termo devem ser iguais. Além disso, conforme Faulstich (2005) e Lyons (1987), os sinônimos existem

apenas quando é possível substituir um termo pelo outro em um mesmo enunciado sem prejudicar o sentido pretendido pelo autor.

O exame das relações originalmente classificadas como sinônimos encontra-se organizado em duas partes. A primeira delas diz respeito àquelas que, após análise, apresentaram-se como relações associativas. A segunda parte refere-se àquelas cuja natureza semântica definia-se como uma relação hierárquica.

5.2.2.1a Relações de natureza associativa

Os seguintes termos relacionados originalmente como sinônimos refletiam, após o exame, relações associativas:

- a) Área contaminada – poluição de solo
- b) Área contaminada – degradação do solo
- c) Área contaminada – contaminação do solo
- d) Custo de saúde – aplicação de recurso na saúde
- e) Custo de saúde – investimento em saúde¹⁷
- f) Investimento em saúde – orçamento da saúde
- g) Violência no trânsito – problema de violência no trânsito

Nos casos *a*, *b*, e *c*, trata-se de uma relação associativa que expressa as noções de **processo/resultado**. Os processos de **<poluição de solo>**, **<degradação do solo>** e **<contaminação do solo>** resultam em **<área contaminada>**.

Os exemplos *d* e *e* foram definidos como relação associativa de oposição, considerando os significados dos elementos **<aplicação>** e **<investimento>** como sendo de sentido oposto ao conceito de **<custo>**.

¹⁷ Registra-se que, para efeito de contagem dos dados do *corpus*, foi considerada também a recíproca desta relação: Investimento em saúde – Custo de saúde

O item *f* foi classificado como uma relação associativa do tipo **coisa/aplicação**, sendo o **<investimento>** a aplicação do **<orçamento>**.

No caso *g*, a relação associativa é do tipo **conceito/atributo**, sendo o conceito de **<violência no trânsito>** relacionado à expressão **<problema de violência no trânsito>** na qual o atributo “problema” é ressaltado.

5.2.2.1b Relações de natureza hierárquica

Os seguintes termos relacionados originalmente como sinônimos foram classificados, após exame, como relações hierárquicas:

- a) Ações de saúde – campanha de saúde¹⁸
- b) Ações de saúde – programas de saúde
- c) Campanha de saúde – programas de saúde
- d) Custo de saúde – orçamento da saúde
- e) Depredação de áreas públicas – depredação de equipamentos públicos
- f) Gangues juvenis – gangues na escola

De forma geral, observa-se nesses exemplos o uso do recurso de deslocamento genérico em que “o nome de uma classe e os nomes de seus membros são tratados como equivalentes em seu significado, ficando o termo mais geral como descritor” (IBICT, 1984 p. 24). Na língua natural, este fenômeno pode ocorrer como uma relação recíproca e simétrica (sinonímica) ou assimétrica (hiponímica), dentro de um mesmo discurso. No entanto, a decisão pelo registro como relação de equivalência em um instrumento de representação do conhecimento pode vir a prejudicar a precisão na recuperação da informação.

¹⁸ Registra-se que, para efeito de contagem dos dados do *corpus*, foi considerada também a recíproca desta relação: Campanha de saúde – ações de saúde

Os casos apresentados em *a*, *c*, *d*, *e* e *f* são de natureza hierárquica, pois se trata do relacionamento entre um termo superordenado (com todas as características) que se liga a termos subordinados (com pelo menos uma característica a mais), conforme Gomes (1990), Cintra (2002), ISO 1087 e ISO 704/2000. Constituem-se, portanto, como uma relação do tipo gênero/espécie, determinadas da seguinte forma em uma representação gráfica de cadeia descendente.

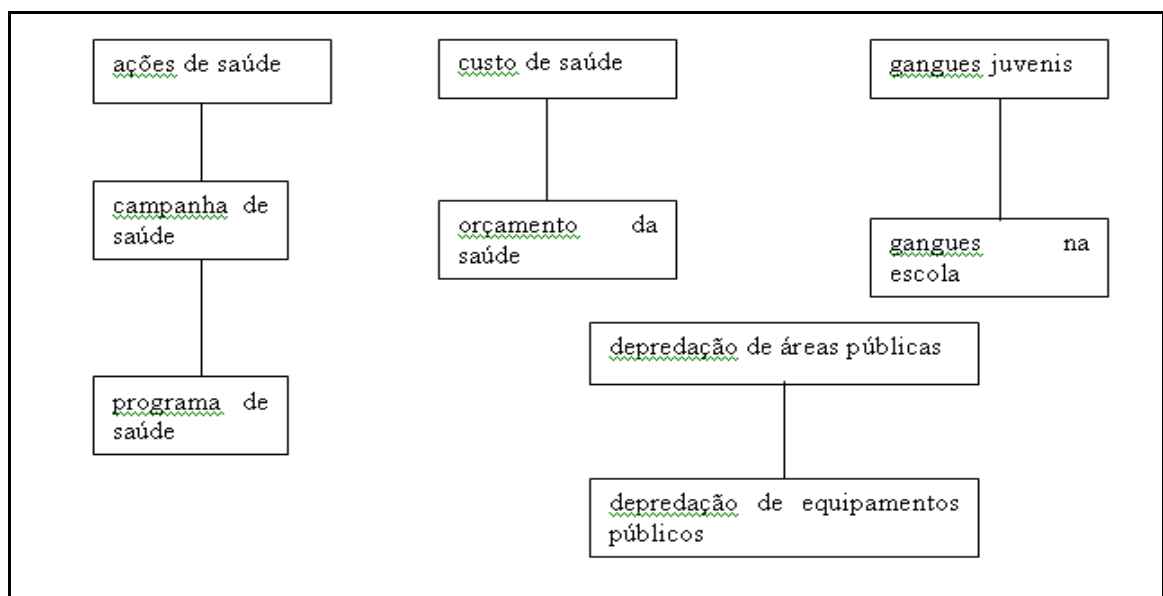


Figura 1: Cadeia descendente das relações gênero/espécie

Quanto ao item *b*, ao interpretá-lo como relação hierárquica e não mais sinonímica, este apresenta um conflito semântico quando estabelece uma relação direta entre <ações de saúde> e <programas de saúde>. Conforme pode ser verificado na Figura 1, entre esses dois termos existe o termo intermediário <campanha de saúde>, caracterizando, portanto, essa relação como indireta e impossível de ser representada diretamente.

No que se refere ao item *f*, ressalta-se ser um exemplo em que se pode dizer que houve um apagamento do elemento central “juvenis” no termo complexo <gângues nas escolas>, tornando o conceito “gângues (juvenis) nas escolas” subentendido na designação. Com base

nessa análise, considera-se a relação como hierárquica em que <ganguês juvenis> possui como termo específico <ganguês (juvenis) na escola>.

5.2.2.2 Relações classificadas originalmente como Hierárquicas

O exame das relações originalmente classificadas como hierárquicas está sistematizado em duas partes. Na primeira, tratam-se os casos em que se concluiu que a natureza semântica entre os termos é do tipo sinonímica. Na segunda parte, encontram-se as relações consideradas como associativas.

5.2.2.2a Relações de natureza sinonímica

Os seguintes termos relacionados originalmente como hierárquicos foram classificados, após a análise, como sinônimos:

- a) Biodiesel – óleo biocombustível ecológico
- b) Combustível biológico – óleo biocombustível ecológico
- c) Prevenção de saúde – prevenção de doença

Tanto os conceitos apresentados no item *a*, *b* como no item *c* se caracterizam como sinônimos, pois eles obedecem ao requisito, definido por Faulstich (2005), de designarem o mesmo conceito e serem intercambiáveis em contextos de igual entendimento.

Ressalta-se que o caso apontado na letra *c* apresenta uma substituição dos elementos <saúde> e <doença> na composição das duas expressões complexas. No entanto, embora sejam conceitos opostos, não é cabível concluir que se trata de uma relação associativa de oposição, pois não pode existir uma ação de prevenção de saúde. Neste caso, provavelmente ocorreu o fenômeno do apagamento do elemento “doença” na expressão <prevenção de saúde> e a substituição da preposição em pela de. É provável que a expressão original fosse

<prevenção (de doença) em saúde>, o que possibilita inferir a natureza sinonímica da relação.

5.2.2.2b Relações de natureza associativa

Os seguintes termos relacionados originalmente como hierárquicos foram identificados, após exame, como associativos:

- a) Prevenção da saúde – educação e prevenção em saúde
- b) Prevenção da saúde – promoção e prevenção em saúde
- c) Prevenção da saúde – serviço de prevenção de saúde
- d) Falta de condição de higiene – origem de doença
- e) Falta de condição de higiene – motivo de doença
- f) Falta de condição de higiene – causa de enfermidades
- g) Falta de condição de higiene – agravo em saúde
- h) Falta de condição de higiene – causa de doenças

Nos exemplos dos itens *a* e *b*, verifica-se uma ampliação conceitual do segundo termo da relação. Mesmo que os exemplos não se enquadrem em nenhum dos parâmetros apresentados por Marroni (2006), decidiu-se incluí-los nas relações associativas, isto por não ser possível a sua identificação com os outros dois tipos de relação semântica.

No item *c*, constata-se uma relação associativa do tipo ação <prevenção da saúde>/ processo <serviço de prevenção de saúde>.

Os itens *d*, *e*, *f*, *g* e *h* caracterizam-se com relação associativa de **causa/efeito**, conforme ilustrado no Quadro 21.

CAUSA	EFEITO
Falta de condição de higiene	origem de doença
Falta de condição de higiene	motivo de doença
Falta de condição de higiene	causa de enfermidades
Falta de condição de higiene	agravo em saúde
Falta de condição de higiene	causa de doenças

Quadro 21: Síntese relação associativa causa/efeito

5.2.2.3 Relações classificadas originalmente como Conexas

O exame das relações descritas originalmente como conexas foram classificadas, após análise, como hierárquicas do tipo **gênero/espécie**. São elas:

- a) Aterro sanitário – aterro sanitário controlado
- b) Aumento de doenças – aumento de enfermidade grave
- c) Aumento de doenças – aumento de enfermidade transmissível

Os pares de expressões descritos pelos itens *a*, *b* e *c* apresentam conceitos que se relacionam de forma verticalizada. Há, em cada par, o compartilhamento de todas as características genéricas, sendo que um dos termos possui, além dessas, uma ou mais características que o especifica em relação ao termo maior.

Após a análise e sugestão de nova classificação dos tipos de relação, os itens que não apresentaram conformidade passaram a ter a seguinte proposta de construção:

Nº	TERMO	RELAÇÃO CONSTRUÍDA	RELAÇÃO PROPOSTA
01	Área contaminada	Poluição do solo Sinônimo	Associativa/conexo Processo / resultado
02	Área contaminada	Degradação do solo Sinônimo	Associativa/conexo Processo / resultado
03	Área contaminada	Contaminação do solo Sinônimo	Associativa/conexo Processo / resultado
04	Ações de saúde	Campanha de saúde Sinônimo	Hierárquica Genérico / específico
05	Ações de saúde	Programas de saúde Sinônimo	Hierárquica indireta Genérico / específico
06	Campanha de saúde	Ações de saúde Sinônimo	Hierárquica Genérico / específico
07	Campanha de saúde	Programas de saúde Sinônimo	Hierárquica Genérico / específico
08	Custo de saúde	Aplicação de recurso na saúde: Sinônimo	Associativa/Conexo oposição
09	Custo de saúde	Orçamento da saúde Sinônimo	Hierárquica Genérico / específico
10	Custo de saúde	Investimento em saúde Sinônimo	Associativa/Conexo oposição
11	Investimento em saúde	Orçamento da saúde Sinônimo	Associativa/Conexo Coisa / aplicação
12	Investimento em saúde	Custo da saúde Sinônimo	Associativa/Conexo oposição
13	Violência no trânsito	Problema da violência no trânsito Sinônimo	Associativa/Conexo Conceito / atributo
14	Gangues juvenis	Gangues nas escolas Sinônimo	Hierárquica Genérico / específico
15	Depredação de áreas públicas	Depredação de equipamentos públicos Sinônimo	Hierárquica Genérico / específico
16	Biodiesel	Óleo biocombustível ecológico Hierárquica	Sinônimo
17	Combustível biológico	Óleo biocombustível ecológico Hierárquica	Sinônimo
18	Prevenção da saúde	Educação e prevenção em saúde Hierárquica	Associativa/Conexo Ação / aplicação
19	Prevenção da saúde	Promoção e prevenção em saúde Hierárquica	Associativa/Conexo Ação / aplicação
20	Prevenção da saúde	Serviço de prevenção de saúde Hierárquica	Associativa/Conexo Ação / processo
21	Prevenção da saúde	Prevenção de doença: Hierárquica	Sinônimo
22	Falta de condição de higiene	Origem de doenças Hierárquica	Associativa/Conexo Causa / efeito
23	Falta de condição de higiene	Motivo de doenças Hierárquica	Associativa/Conexo Causa / efeito
24	Falta de condição de higiene	Causa de enfermidades Hierárquica	Associativa/Conexo Causa / efeito
25	Falta de condição de higiene	Agravo em saúde Hierárquica	Associativa/Conexo Causa / efeito
26	Falta de condição de higiene	Causa de doenças Hierárquica	Associativa/Conexo Causa / efeito
27	Aterro sanitário	Aterro sanitário controlado Conexo	Hierárquica Genérico / específico
28	Aumento de doenças	Aumento de enfermidade grave Conexo	Hierárquica Genérico / específico
29	Aumento de doenças	Aumento de enfermidades transmissíveis: Conexo	Hierárquica Genérico / específico

Quadro 22: Relações construídas e relações propostas

Esta nova classificação das relações tende a contribuir na qualidade das informações recuperadas no sistema. Por isso, com o objetivo de verificar o comportamento do sistema ante a classificação proposta, procedeu-se a alteração da relação dos termos “gângues juvenis” e “gângues nas escolas” originalmente classificadas como sinônimos para **hierárquica**.

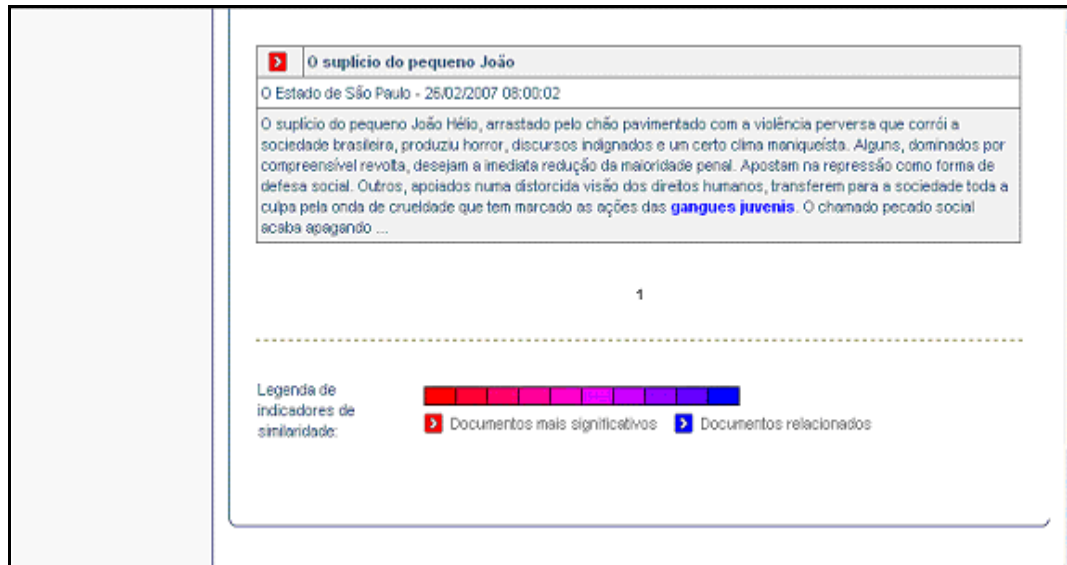


Figura 2: Alteração da relação dos termos “gângues juvenis” e “gângues nas escolas”

No que se refere à quantidade de documentos recuperados, foi possível verificar que a classificação original resultou em 36 documentos, o que apresenta um conteúdo mais genérico. Após a alteração proposta, a recuperação resultou em apenas um documento de conteúdo mais específico. Considerando a sequência de peso da métrica de similaridade, observou-se que, além da quantidade, também a ordem dos documentos recuperados pelo sistema sofreu alteração, porém ficou mantida a pertinência e contextualização do primeiro ao último documento recuperado.

CAPÍTULO 6 - CONSIDERAÇÕES FINAIS

A dinamicidade tecnológica disponível atualmente impõe uma nova realidade a todas as áreas do conhecimento. Nesse contexto, observou-se uma interdisciplinaridade crescente aliada à necessidade recorrente de descobertas. Por outro lado, o meio acadêmico e profissional reconhece que, devido à sua imensurável quantidade, as informações disponíveis, por certo, jamais serão acessadas e absorvidas na sua totalidade, pois, ao mesmo tempo, mais e mais conteúdos são gerados a todo o momento. Como consequência, os Sistemas de Recuperação de Informações vêm assumindo responsabilidades cada vez maiores, arriscando-se dizer que eles se tornaram até imprescindíveis.

Entretanto, a representação do conhecimento, enquanto parte essencial dos Sistemas, precisa ser sistematizada, tratada, armazenada e acessada dentro de uma contextualização representativa. Nesse sentido, verificou-se que a concepção de um Sistema envolve e exige diversos fatores, como agentes, objetivos e etapas que buscam a identificação do melhor escopo e da melhor resposta. Essas etapas culminaram na verificação das funcionalidades da ferramenta desenvolvida, na expectativa de que os resultados obtidos fossem os resultados esperados quando da sua concepção. Cumpre ressaltar que o Sistema KMAI®, descrito no estudo, utilizou a “Engenharia da Mente”, proposta por Bueno (2005), como metodologia para representação do conhecimento.

Neste estudo, cujo objetivo se voltou à validação das ontologias construídas e inseridas em um Sistema, a fase de testes envolveu os recursos tecnológicos da ferramenta KMAI®, que permitiu a verificação das funcionalidades a que se propõe no Sistema ORBIS_MC. Os testes das ontologias foram feitos diretamente na ferramenta, em que cada termo escolhido foi alvo de busca nas fontes de informação, previamente definida, e que

formam a base referencial do Sistema. Buscou-se, nestes testes, verificar se as ontologias têm representatividade, se a forma de construção está adequada, se a linguagem está de acordo com os textos-base e com os usuários e, o item mais importante, se as relações construídas traduzem o conhecimento alvo de representação.

Constatou-se que essas funcionalidades, que utilizam o recurso de ontologias como forma de representação e recuperação do conhecimento, possibilitaram a validação dos termos ou expressões e suas respectivas relações semânticas. Diante disso, pode-se dizer que a validação está diretamente relacionada à qualidade da recuperação de informações, no que tange à contextualização e expectativas do usuário no momento da busca.

Foram identificados também os parâmetros que nortearam os trabalhos das equipes diretamente envolvidas na construção de vocabulários. Entretanto, na fase de testes de construção de vocabulários, etapa considerada fundamental, pelo fato do processo ser altamente subjetivo e pouco estudado, sentiu-se a falta de uma metodologia consolidada, de cunho lingüístico, de validação de ontologias. Por isso, o estudo do caso proposto demonstrou ser possível a implementação de critérios lingüísticos para a fase de validação dos termos, além de reforçar sua importância. Uma construção coerente e bem elaborada evita, por exemplo, que sejam identificadas relações que gerem uma recuperação sem pertinência ao assunto pesquisado. Para minimizar esse problema, verificou-se a necessidade do conhecimento prévio do contexto assim como o entendimento deste de forma sincronizada entre todos os envolvidos.

Assim, a elaboração de uma metodologia que contenha parâmetros lingüísticos poderá contribuir na construção das relações semânticas, favorecendo o entendimento e a inserção das expressões identificadas dentro do contexto do domínio de aplicação. Apoiado nos resultados encontrados, entende-se que a contribuição deste trabalho está na definição e sistematização de critérios semânticos que possam nortear a capacitação, a construção e os

testes de validação, garantindo uma melhor performance do sistema, além de uma melhor organização da base de conhecimento. Esta organização vai refletir diretamente no comportamento do usuário frente à ferramenta, garantindo adequação do texto de entrada à verificação dos documentos existentes.

No caso do Sistema KMAI® ORBIS, constatou-se um total de 29 inconsistências num universo estudado de 163. Concluiu-se ser um índice baixo, considerando que não faz parte da metodologia de capacitação para a construção o módulo específico para o conhecimento e entendimento da correspondência lingüística. Também não faz parte da metodologia de testes do sistema a etapa específica de avaliação da ontologia sob o aspecto lingüístico. Contudo, há de se ressaltar que a equipe responsável pela construção do vocabulário foi coordenada por especialista com profundo conhecimento na área, que participou ativamente de todas as etapas do desenvolvimento do sistema e da construção das ontologias. Esta, porém, é uma situação ímpar que não pode ser considerada regra para todos os sistemas.

Neste esteio, sugere-se que os parâmetros e critérios apresentados neste trabalho sejam sistematizados de forma a servir como fonte de informação de apoio à capacitação em Construção de Ontologias. Propõe-se ainda que este material faça parte das rotinas de testes com objetivo de garantir a qualidade lingüística das expressões.

Como forma de contribuir para a qualidade das ontologias construídas, sugere-se que seja implementado, na capacitação de Engenharia de Ontologias, um módulo de conhecimento dos critérios lingüísticos apresentados, em que:

- a) Para construção de sinônimos, sejam considerados termos que têm o mesmo conceito, sendo intercambiáveis em todos os contextos;
- b) Para construção de relações hierárquicas, sejam considerados termos superordenados (com todas as características) que se liguem a termos subordinados (com pelo menos

uma característica a mais), caracterizando o tipo genérico/específico; também que sejam considerados termos que apresentem relação partitiva (o todo e suas partes);

- c) Para construção de relações associativas (conexas), sejam considerados os dez parâmetros propostos, respeitando a ligação temática entre os termos.

Por último, considerando que a nova geração de ferramentas tecnológicas se apóia nos recursos da Inteligência Artificial, que buscam se tornar cada vez mais inteligentes, as ontologias se apresentam como uma forma dinâmica de representação do conhecimento em que a contextualização dos conteúdos, durante o processo de construção, é a mesma quando do processo de recuperação. Portanto, pode-se dizer que a representação do conhecimento de um domínio pode ser aplicada a qualquer área do conhecimento, sendo imprescindível para construção e utilização de ferramenta de recuperação, seja ela manual ou eletrônica.

Para futuros trabalhos, sugere-se aliar esta validação, norteada por critérios lingüísticos, à metodologia de testes de recuperação, utilizando as ontologias construídas no desenvolvimento dos sistemas.

REFERENCIAL BIBLIOGRÁFICO

BARDIN, Laurence. **Análise de conteúdo**. Lisboa: Edições 70, 2004.

BUENO, Tania Cristina D'Agostini. Engenharia da Mente: uma metodologia de representação do conhecimento para construção de ontologias em sistemas baseados em conhecimento. Florianópolis: UFSC. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, 2005.

CABRÉ, M. Teresa. **La terminología: teoría, metodología, aplicaciones**. Barcelona: Editorial Antártica/Empúries, 1993.

CÂMARA Jr., J. Mattoso. **Dicionário de lingüística e gramática**. 17. ed. Petrópolis: Vozes, 1996.

CÂMARA, J.C. Learning Metadata Standards. Barcelona, 2002. Disponível em: <<http://www.iau.upf.es/~jblat/material/doctorat/students/jccbis/Ontologias.htm>>. Acesso em 13 nov. 2005.

CHANDRASEKARAN, B., *et al.* What Are Ontologies, and Why Do We Need Them?. Amsterdam, 1999. Disponível em: <<http://www.cs.umbc.edu/771/papers/chandrasekaranetal99.pdf>>. Acesso em mai 2006.

CINTRA, Anna Maria Marques, *et al.* **Para entender as linguagens documentárias**. 2. ed. rev. e ampl. São Paulo: Polis, 2002.

DAHLBERG, I. Teoria do conceito. Ciência da Informação, Brasília, v. 7, n. 2, p. 101-107, 1978.

DESCARTES, René. **Discurso do Método**. Coleção Grandes Obras do Pensamento Universal. São Paulo, Escala: s.d.

DICTIONARY.COM, 2004. Disponível em <<http://www.dictionary.com>>

DUBOIS, Jean; *et al.* **Dicionário de Lingüística**. São Paulo: Cultrix, 1986.

FALBO, Ricardo A., *et al.* Ontologias e ambientes de desenvolvimento de software semânticos. **Actas de las IV Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento**, IIISIC'2004. Volumen I, pp. 277-292. Madri, 2004. Disponível em <<http://www.inf.ufes.br/%7Efalbo/download/pub/2004-IIISIC-1.pdf>>. Acesso em 07 mai 2006.

FAULTSTICH, Emilde. **Como ler, entender e redigir um texto**. 18. ed. São Paulo: Vozes, 2005.

FEITOSA, A. L. G. A integração entre sistemas legislativos, terminologia e web semântica na organização e representação da informação legislativa. Brasília: UnB. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, 2005.

FEITOSA, Ailton. **Organização da informação na Web: das tags à web semântica**. Brasília: Thesaurus, 2006.

FELBER, H.; PICHT, H. **Metodos de terminografia y principios de investigacion terminologica**. Madrid: Instituto “Miguel de Cervantes”.CSIC: 1984.

FELBER, Helmut. **Terminology Manual**. Paris: Unisist, 1984.

FLICK, Uwe. **Uma introdução à Pesquisa Qualitativa**. 2. ed. Porto Alegre: Bookman, 2004.

Glossário de Gramática e Lingüística, disponível em <http://www.paulohernandes.pro.br/glossario/indice.html>, acesso em 01 fev 2006.

GOLDENBERG, Mirian. **A arte de pesquisar: como fazer pesquisa qualitativa em Ciências Sociais**. Rio de Janeiro: Record, 2004.

GOMES, H. E. (coord.). **Manual de elaboração de tesaurus monolíngües**. Brasília: Ministério da Educação/Ministério da Ciência e Tecnologia, Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.

GOMÉZ-PÉREZ, A. Ontological engineering: a state of the art. Expert Update. 1999.

Disponível em <

<http://citeseer.ist.psu.edu/cache/papers/cs/22343/http:zSzzSzwww.csc.liv.ac.ukzSz~franzSzExpertUpdatezSzontologies.pdf/ontological-engineering-a-state.pdf>> acesso em 06 mai 2006.

GOUADEC, Daniel. **Terminologie: Constitution des données**. Paris: Afnor, 1990.

Gramática da Língua Portuguesa, disponível em

<<http://www.portugues.com.br/semantica/homonimia.asp>> acesso em 01 fev 2006.

GRUNINGER, Michael, MARK S. Fox, Methodology for the Design and Evaluation of Ontologies, **Proceedings of IJCAI'95**, Workshop on Basic Ontological Issues in Knowledge Sharing, April, 1995. Disponível em

<www.isye.gatech.edu/~lfm/8851/Sources/Ontology/FoxOntologies.ps>. Acesso em 13 abr 2006.

GRUBER, Thomas R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *In: International Journal Human-Computer Studies* 43, p.907-928. Padova, 1993.

Disponível em < <http://www.cise.ufl.edu/~jhammer/classes/6930/XML-FA02/papers/gruber93ontology.pdf>>. Acesso em 01 mai 2006.

GRUBER, Tom. What is an Ontology? Disponível em < <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>.

Acesso em 18 out 2005.

GUARINO, Nicola. Formal Ontology and Information Systems. *In: Proceedings of FOIS'98* p.3-15. Italy, 1998. Disponível em< <http://www.loa-cnr.it/Papers/FOIS98.pdf>>. Acesso em 16 mai 2006.

GUILLÉN, A. I. S.; MORENO, J. M. P.; BADAYA, D. M. Ontologías para la gestión del conocimiento. Disponível em: < <http://es.geocities.com/ontologia04/index.htm>>. Acesso em 13 nov. 2005.

GUIMARÃES, Francisco José Zamith. Utilização de ontologias no domínio B2C. Rio de Janeiro: PUC. Dissertação (Mestrado em Informática) – Pontifícia Universidade Católica, 2002.

GUINCHAT, Claire; MENOU, Michel. **Introdução geral às ciências e técnicas da informação e documentação**. Brasília: IBICT, 1994.

HYVÖNEN, E. The Semantic Web – The new Internet of Meanings. *In: HYVÖNEN, Eero* (editor). **Semantic Web Kick-Off in Finland: vision, technologies, research and applications**. Helsinki, Finland: HIIT Publications. 2002. Disponível em: <

<http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/kick-off/proceedings.pdf>>. Acesso em: 06 mai. 2006.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA.
Diretrizes para elaboração de tesauros monolíngües. Brasília: 1984.

INTERNATIONAL STANDARDS ORGANIZATION. *ISO 704:2000: Terminology work – Principles and methods.* [S.l.], 2000.

INTERNATIONAL STANDARDS ORGANIZATION. *ISO/R 1087-1969: Vocabulary of terminology.* [S.l.], 1982.

INTERNATIONAL STANDARDS ORGANIZATION. *ISO/R 2788-1974: Documentation – Guidelines for the establishment and development of monolingual thesauri.* [S.l.], 1982.

KRIEGER, M. da G.; FINATTO, M.J.B. **Introdução à terminologia:** teoria & prática. São Paulo: Contexto, 2004.

LANCASTER, F.W. **Indexação e resumos:** teoria e prática. Brasília: Briquet de Lemos, 1997.

LARA, Marilda Lopez Ginez de. Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. *Ciência da Informação*. Brasília, v. 33, n.2, p. 91-96, maio/ago 2004.

Le COADIC, Yves-François. **A ciência da Informação.** 2. ed. Brasília: Briquet de Lemos, 2004.

LÉVY, Pierre. **O que é virtual.** São Paulo: Editora 34, 1998.

LIMA-MARQUES, Mamede. **Ontologias da filosofia à representação do conhecimento.** Brasília: Thesaurus, 2006.

LYONS, J. **Linguagem e lingüística: uma introdução.** Rio de Janeiro: LTC, 1987.

MARRONI, Gilza Núria Brandão. Identificação e delimitação de relações associativas em tesauros: um estudo de caso na área do direito do trabalho. Brasília: UnB. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, 2006.

McGARRY, Kevin. **O contexto dinâmico da informação.** Brasília: Briquet de Lemos, 1999.

MOREIRA, Alexandra; OLIVEIRA, Alcione de Paiva. Contribuição da terminologia na modelagem de sistemas computacionais. *DataGramaZero*, Brasília, v.6, n.5, out./05.

NETIC. Núcleo de estudos em tecnologias para informação e conhecimento. 2003. Gestão de Conteúdo web através de ontologias: conceitos e aplicações. Disponível em: <<http://www.fernando.parreiras.nom.br/palestras/infoimagem.pdf>>. Acesso em 13 nov. 2005.

Observatório Regional Base de Indicadores de Sustentabilidade Metropolitano de Curitiba – ORBIS_MC. Disponível em <http://www.observatorio.com.br>.

PARREIRAS, F. Introdução à construção de ontologias. Disponível em: <<http://www.fernando.parreiras.nom.br/palestras/engeonto.pdf>>. Acesso em 13 nov. 2005.

SAGER, Juan C. **A practical Course in terminology processing.** Philadelphia: John Benjamins, 1990.

SALOMON, Délcio Vieira. **A maravilhosa incerteza:** ensaio de metodologia dialética sobre a problematização no processo do pensar, pesquisar e criar. São Paulo, Martins Fontes: 2000.

SANTOS, Carlos Almeida; SARMENTO, Manoel Soares Filho. Linguagens de especialidade: um inventário de verbetes para dicionários técnico-científicos. In: **IX Congresso Nacional de Linguística e Filologia**, 2005, Rio de Janeiro. Caderno da

Programação do Evento. Rio de Janeiro : UERJ, 2005. v. 1. p. 16-16. Disponível em < <http://www.filologia.org.br/ixcnlf/16/10.htm>>, acesso em 19 mar 2006.

SAUSSURE, Ferdinand de. **Curso de lingüística geral**. 25. ed. São Paulo: Cultrix, 2003.

SOWA, John F. Building, Sharing, and Merging Ontologies. Disponível em <http://users.bestweb.net/~sowa/ontology/ontoshar.htm>. Acesso em 31 abr 2006.

SWARTOUT, Bill, *et al.* Toward Distributed Use of Large-Scale Ontologies. Disponível em < http://ksi.cpsc.ucalgary.ca/KAW/KAW96/swartout/Banff_96_final_2.html>. Acesso em 20 maio 2006.

TELLO, Adolfo Lozano. Ontologías em la Web Semántica. *In: Iº Jornadas de Ingeniería Web'01*. Cáceres, xxxx. Disponível em < <http://www.informandote.com/jornadasIngWEB/articulos/jiw02.pdf>>. Acesso em 10 set 2006.

TOBAR, Federico. Yalour, YALOUR, Margot Romano. **Como fazer teses em saúde pública**. Rio de Janeiro, Fiocruz:2004.

USCHOLD, Mike. Building Ontologies: towards a unified methodology. *In: Proc. Expert Systems 96*, Cambridge, December 16-18 . Disponível em <<http://citeseer.ist.psu.edu/uschold96building.html>>, acesso em 07 mai 2006.

ANEXOS

SINONÍMIA

DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: ENERGIA			
TERMO	RELAÇÃO	ADEQUAÇÃO	COMENTÁRIO
Biodiesel	Combustível natural	ok	
	Combustível biológico	ok	
	Biocombustível	ok	
Combustível biológico	Biodiesel	ok	
	Combustível natural	ok	
	Biocombustível	ok	
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: GESTÃO DE RECURSOS NATURAIS			
Áreas Verdes	Zonas verdes	ok	
	Espaços verdes	ok	
Manejo Ambiental	Manejo de recurso natural	ok	
	Gestão do Ambiente e de seus recursos	ok	
	Gerenciamento de recursos naturais	ok	
	Gestão de recurso natural	ok	
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: SANEAMENTO AMBIENTAL			
Aterro sanitário	Depósitos de resíduos sólidos urbanos	ok	
	Depósitos de lixo	ok	
Coleta e Tratamento de resíduos	Tratamento de resíduos	ok	
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: POLUIÇÃO AMBIENTAL			
Área contaminada	Poluição do solo	não	Associativa Processo/resultado
	Degradação do solo	não	Associativa Processo/resultado
	Contaminação do solo	não	Associativa Processo/resultado
Destruição de Recursos Hídricos	Poluição hídrica	ok	
	Degradação de recursos hídricos	ok	
	Degradação dos rios	ok	
	Deterioração dos rios	ok	
	Destruição dos rios	ok	
	Deterioração de recursos hídricos	ok	
	Contaminação de recursos hídricos	ok	
	Poluição de recursos hídricos	ok	
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: OCUPAÇÃO DO SOLO			
Explosão demográfica urbana	Crescimento da população	ok	
	Expansão populacional	ok	

	Aumento populacional nas cidades	ok	
	Crescimento populacional	ok	
Loteamento irregular	Loteamento clandestino	ok	
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: ÓRGÃOS VINCULADOS A SAÚDE AMBIENTAL Nenhuma relação encontrada			
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: PROMOÇÃO DA SAÚDE			
Ações de educação em saúde	Práticas de educação em saúde	ok	
	Atividades de educação em saúde	ok	
	Educação em saúde	ok	
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: VIGILÂNCIA EM SAÚDE			
Aumento de doenças	Casos de aumento de enfermidades	ok	
	Surto de doenças	ok	
	Surto de epidemias	ok	
	Aumento de epidemias	ok	
	Aumento de enfermidades	ok	
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: POLÍTICAS DE SAÚDE			
Ações de saúde	Campanha de saúde	nao	Hierarquica generico/especifico
	Programas de saúde	não	Hierarquica generico/especifico
Campanha de saúde	Ações de saúde	não	Hierarquica generico/especifico
	Programas de saúde	não	Hierarquica generico/especifico
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: GESTÃO DE SAÚDE			
Custo de saúde	Aplicação de recurso na saúde	não	Associativa oposição
	Orçamento da saúde	não	Hierárquica genérico/específico
	Investimento em saúde	não	Associativa oposição
Investimento em saúde	Orçamento da saúde	não	Associativa coisa/aplicação
	Custo da saúde	não	Associativa oposição
	Aplicação de recurso na saúde	ok	
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: MODALIDADES DE VIOLÊNCIA URBANA			
Violência nos presídios	Violência prisional	ok	
Violência no trânsito	Problema da violência no trânsito	não	Conexo Conceito/atributo
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: VIOLÊNCIA BASEADA POR LOCALIDADE			
Violência doméstica	Violência no lar	ok	
	Violência familiar	ok	
	Agressão doméstica	ok	
	Violência domiciliar	ok	
	Agressão familiar	ok	
Violência na periferia	Violência no bairro	ok	

	violência na comunidade	ok	
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: CAUSAS DA VIOLÊNCIA URBANA			
Gangues juvenis	Gangues nas escolas	nao	Hierararquica –generico/especifico
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIOS: CONSEQUÊNCIAS DA VIOLÊNCIA URBANA			
Depredação de áreas públicas	Depredação de equipamentos públicos	não	Hierárquica –genérico/específico
	Destruição das áreas públicas	ok	
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: SEGURANÇA PÚBLICA			
Custos da violência	Custos da criminalidade	ok	
	Custos econômicos da violência	ok	
Policiamento comunitário	Parceria da comunidade com a polícia	ok	
	Participação da comunidade na segurança	ok	

HIERÁRQUICA

DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: ENERGIA			
TERMO	RELAÇÃO	ADEQUAÇÃO	COMENTÁRIO
Biodiesel	Óleo biocombustível ecológico	não	sinonimo
Combustível biológico	Óleo biocombustível ecológico	não	sinonimo
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: GESTÃO DE RECURSOS NATURAIS			
Áreas verdes	Espaços verdes integrados	ok	
	Rede de espaços verdes	ok	
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: SANEAMENTO AMBIENTAL			
Coleta e Tratamento de resíduos	Tratamento de resíduos químicos	ok	Genérico/específico
	Tratamento de resíduos sanitários	ok	Genérico/específico
	Tratamento de resíduos sólidos	ok	Genérico/específico
	Tratamento de águas residuais	ok	Genérico/específico
	Tratamento de resíduos industriais	ok	Genérico/específico
	Tratamento de resíduos orgânicos	ok	Genérico/específico
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: PROMOÇÃO DA SAÚDE			
Prevenção da saúde	Educação e prevenção em saúde	não	Associativa
	Promoção e prevenção em saúde	não	Associativa

	Serviço de prevenção de saúde	não	Associativa ação/processo
	Programa de saúde pública	ok	
	Prevenção de doença	não	Sinônimo
	Questão de saúde	ok	
	Questão de saúde pública	ok	
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: VIGILÂNCIA EM SAÚDE			
Falta de condição de higiene	Origem de doenças	não	Associativa Causa/efeito
	Motivo de doenças	não	Associativa causa/efeito
	Causa de enfermidades	não	Associativa causa/efeito
	Agravo em saúde	não	Associativa causa / efeito
	Causa de doenças	não	Associativa causa/efeito
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: POLÍTICAS DE SAÚDE			
Ações de saúde	Ações de saúde coletiva	ok	Genérico/específico
	Ações de saúde pública	ok	Genérico/específico
Campanha de saúde	Ações de saúde coletiva	ok	Genérico/específico
	Ações de saúde pública	ok	Genérico/específico
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: MODALIDADES DE VIOLÊNCIA URBANA			
Violência nos presídios	Tortura de detentos	ok	Genérico/específico
	Maus-tratos aos detentos	ok	Genérico/específico
Violência no trânsito	Roubo no trânsito	ok	Genérico/específico
	Seqüestros no trânsito	ok	Genérico/específico
	Homicídios no trânsito	ok	Genérico/específico
	Assassinatos no trânsito	ok	Genérico/específico
	Assalto no trânsito	ok	Genérico/específico
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: CAUSAS DA VIOLÊNCIA URBANA			
Gangues juvenis	Adolescentes infratores	ok	
	Menores infratores	ok	

ASSOCIATIVAS (conexos)

DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: ENERGIA			
TERMO	RELAÇÃO	ADEQUAÇÃO	COMENTÁRIO
Biodiesel	Pesquisa sobre biocombustível	ok	coisa/processo
	Aumento da produção de biodiesel	ok	coisa/resultado
	Comercialização de biodiesel	ok	coisa/processo
Combustível biológico	Pesquisa sobre biocombustível	ok	coisa/processo
	Aumento da produção de biodiesel	ok	coisa/aplicação
	Comercialização de biodiesel	ok	materia prima/resultado
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: GESTÃO DE RECURSOS NATURAIS			
Áreas verdes	Recuperação de espaços verdes	ok	coisa/processo

	Espaços verdes de proteção	ok	matéria prima/processo
	Gestão de espaços verdes	ok	matéria prima/processo
	Regulamentação de espaços verdes	ok	coisa/processo
Manejo ambiental	Produção de recursos naturais	ok	ação/resultado
	Políticas de gestão de recursos naturais	ok	Ação/produto
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: SANEAMENTO AMBIENTAL			
Aterro sanitário	Fechamento de aterros sanitários	ok	coisa/processo
	Saturamento do aterro sanitário	ok	coisa/resultado
	Aterro sanitário controlado	não	Hierárquico genérico/específico
Coleta e tratamento de resíduos	Empresas de tratamento de resíduos	ok	Agente/processo
	Central de tratamento de resíduos	ok	Agente/processo
	Sistemas de tratamento de resíduos	ok	Agente/processo
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: POLUIÇÃO AMBIENTAL			
Área contaminada	Grau de contaminação do solo	ok	campo estudo/fenomeno estudado
	Situação de contaminação do solo	ok	campo de estudo/objeto estudado
	Nível de contaminação do solo	ok	campo de estudo/objeto ou fenômeno estudado
	Processo de contaminação do solo	ok	causa/objeto ou fenômeno estudado
	Aumento da degradação do solo	ok	causa/efeito
	Redução da degradação do solo	ok	oposição
	Controle da degradação do solo	ok	coisa/processo
Coleta dos Recursos Hídricos	Eutrofização de lagos	ok	oposição
	Salinização de rios e açudes	ok	Causa/efeito
	Doenças ligadas à poluição hídrica	ok	Causa/efeito
DOMÍNIO: SAÚDE AMBIENTAL SUBDOMÍNIO: OCUPAÇÃO DO SOLO			
Explosão demográfica urbana	Crescimento populacional desordenado	ok	ação/resultado
Loteamento irregular	Regularização de loteamentos irregulares	ok	ação/processo
	Regularização de loteamentos clandestinos	ok	ação/processo
	Proliferação de loteamentos clandestinos	ok	causa/efeito
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: VIGILÂNCIA EM SAÚDE			

Aumento de doenças	Aumento de enfermidade grave	não	Hierárquica Genérico/específico
	Aumento de enfermidades transmissíveis	não	Hierárquica Genérico/específico
Falta de condição de higiene	Falta de saneamento	ok	ação/agente
	Condições de hygiene	ok	ação/alvo
	Falta de saneamento básico	ok	ação/agente
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: POLÍTICAS DE SAÚDE			
Ações de saúde	Implantação de programa de saúde	ok	ação/agente
	Falta de programa de saúde	ok	oposição
	Execução de programa de saúde	ok	ação/processo ou aplicação
	Planejamento de programa de saúde	ok	ação/processo ou aplicação
Campanha de saúde	Falta de programa de saúde	ok	oposição
	Execução de programa de saúde	ok	ação/processo ou aplicação
	Planejamento de programa de saúde	ok	ação/processo ou aplicação
	Implantação de programa de saúde	ok	ação/processo ou aplicação
	Campanha de saúde pública	ok	ação/processo ou aplicação
DOMÍNIO: SAÚDE HUMANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: GESTÃO DE SAÚDE			
Custo de saúde	Aumento do investimento em saúde	ok	conceito/mecanismo de medida
	Redução do investimento em saúde	ok	conceito/mecanismo de medida
	Recuperação de investimento em saúde	ok	causa/efeito
Investimento em saúde	Aumento do investimento em saúde	ok	conceito/mecanismo de medida
	Redução do investimento em saúde	ok	conceito/mecanismo de medida
	Recuperação de investimento em saúde	ok	ação/alvo
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: MODALIDADES DE VIOLÊNCIA URBANA			
Violência nos presídios	Rebeliões prisionais	ok	causa/origem
	Aumento da violência nos presídios	ok	conceito/mecanismo de medida
Violência no trânsito	Combate à violência no trânsito	ok	Conceito/alvo
DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA SUBDOMÍNIO: CAUSAS DA VIOLÊNCIA URBANA			
Gangues juvenis	Vandalismo nas escolas	ok	causa/efeito
Tráfico de armas	Tiroteio entre traficantes	ok	causa/efeito
	Controle do tráfico de armas	ok	ação/alvo
	Combate ao tráfico de armas	ok	ação/alvo
	Tiroteio entre quadrilhas	ok	causa/efeito

DOMÍNIO: VIOLÊNCIA URBANA NA REGIÃO METROPOLITANA DE CURITIBA				
SUB DOMÍNIOS: CONSEQUÊNCIAS DA VIOLÊNCIA URBANA				
Mortalidade homicídio	por	Aumento de homicídios	ok	Conceito/mecanismo de medida
		Aumento do número de homicídios	ok	Conceito/mecanismo de medida
		Redução do número de homicídios	ok	Conceito/mecanismo de medida